

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2001, 09:29:30 ; Search time 60.99 Seconds  
(without alignments)  
119.849 Million cell updates/sec

Title: US-09-485-951-2  
Perfect score: 1917  
Sequence: 1 MAFSGSQAPYLSPAVPFSCT.....LPTINRLEVGDIQLTHVQT 355

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Archived: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1633	85.2	311	3	US-08-946-914-4
2	783	40.8	149	2	US-08-788-584-3
3	661	34.5	145	2	US-08-788-584-1
4	603.5	31.5	324	3	US-08-946-914-11
5	574	29.9	323	1	US-08-469-667-16
6	574	29.9	323	3	US-08-946-914-2
7	574	29.9	323	5	PCT-US95-07289-16
8	535	27.9	145	2	US-08-788-584-5
9	535	27.9	145	3	US-08-946-914-12
10	472.5	24.6	316	4	US-09-131-648-5
11	471.5	24.6	317	3	US-08-946-914-6
12	454.5	23.7	316	2	US-08-728-521-3
13	454.5	23.7	316	2	US-08-647-960-2
14	454.5	23.7	316	3	US-08-946-914-15
15	454.5	23.7	316	3	US-08-946-914-17
16	353.5	18.4	264	2	US-08-728-521-1
17	328.5	17.1	264	1	US-08-562-311-4
18	321.5	16.8	336	4	US-09-131-648-1
19	321	16.7	262	3	US-08-946-914-14
20	318.5	16.6	250	3	US-08-946-914-10
21	317.5	16.3	250	1	US-08-562-311-2
22	255	13.3	200	3	US-08-946-914-8
23	254.5	13.3	177	2	US-08-647-960-6
24	250	13.0	147	2	US-08-647-960-7
25	230.5	12.0	136	3	US-08-946-914-13
26	226.5	11.8	135	2	US-08-647-960-5
27	198.5	10.4	132	2	US-08-647-960-11

28	189	9.9	131	2	US-08-647-960-9	Sequence 9, Appl
29	184	9.6	146	2	US-08-647-960-8	Sequence 8, Appl
30	179	9.3	184	2	US-08-647-960-10	Sequence 10, Appl
31	167	8.7	135	3	US-08-050-259B-18	Sequence 18, Appl
32	154	8.0	135	3	US-08-946-914-16	Sequence 16, Appl
33	154	8.0	135	3	US-08-050-259B-20	Sequence 20, Appl
34	154	8.0	135	4	US-09-489-292-2	Sequence 2, Appl
35	142.5	7.4	142	6	5242807-2	Patent No. 5242807
36	129	6.7	132	3	US-08-946-914-9	Sequence 9, Appl
37	115	6.0	402	6	5352575-5	Patent No. 5352575
38	112	5.8	135	2	US-08-647-960-3	Sequence 3, Appl
39	108.5	5.7	135	3	US-08-050-259B-16	Sequence 16, Appl
40	107	5.6	38	1	US-08-540-202-2	Sequence 2, Appl
41	105.5	5.5	366	6	5470718-4	Patent No. 5470718
42	96.5	5.0	511	1	US-08-220-151-17	Sequence 17, Appl
43	96.5	5.0	511	1	US-08-413-118-17	Sequence 17, Appl
44	96.5	5.0	511	3	US-08-473-446-17	Sequence 17, Appl
45	94.5	4.9	328	3	US-09-253-682-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1  
US-08-946-914-4  
; Sequence 4, Application US/08946914  
; Patent No. 6027916  
; GENERAL INFORMATION:  
; APPLICANT: Ni, Jian  
; APPLICANT: Gentz, Reiner L.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
; STREET: 1100 New York Ave., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/946,914  
; FILING DATE: Herewith  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/028,093  
; FILING DATE: 09-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2500  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 311 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-946-914-4

Query Match 85.2%; Score 1633; DB 3; Length 311;  
Best Local Similarity 87.6%; Pred. No. 3.2e-164;  
Matches 311; Conservative 0; Mismatches 0; Indels 44; Gaps 1;  
OY 1 MAFSGSQAPYLSPAVPFSCTIQGGLODGLQITVNGTVLSSSGTREFAVNFQTFSGNDIAF 60

Db 1 MAFSGSQAPYLSAPVPSFSGTIOGLDGLQITVNGTSLSSGTRFAVNFQTFSGNDIAF 60  
QY 61 HFNPRFEDGGYVVCNTRONGSWGPEERKTHMPFKGMPDPDLCLFVLOSSDFKVMVNGILFV 120  
Db 61 HFNPRFEDGGYVVCNTRONGSWGPEERKTHMPFKGMPDPDLCLFVLOSSDFKVMVNGILFV 120  
QY 121 QYFHRVPFHRVDTSVNGSVQLSYISFQNPRTVPVQFAFSTVPFSQPVCFPPRPRGRQK 180  
Db 121 QYFHRVPFHRVDTSVNGSVQLSYISFQ----- 148  
QY 181 PGWTFANPAPITQTVIHTVQSAPGOMFSTPAIPPMYHPAYPMPFITTILGGLYPSKS 240  
Db 149 -----TQTVIHTVQSAPGOMFSTPAIPPMYHPAYPMPFITTILGGLYPSKS 196  
QY 241 ILLSGTVLPQAORFHNLCSGNHIAFHLPNPRFDENAVVRNTQIDNSWGSEERSLPRKMPF 300  
Db 197 ILLSGTVLPQAORFHNLCSGNHIAFHLPNPRFDENAVVRNTQIDNSWGSEERSLPRKMPF 256  
QY 301 VRQGSFVWLCEAHCLKVAVDGQHLFEYHRLNRLPTINRLEVGGDIQLTHVQT 355  
Db 257 VRQGSFVWLCEAHCLKVAVDGQHLFEYHRLNRLPTINRLEVGGDIQLTHVQT 311

## RESULT 2

US-08-788-584-3  
; Sequence 3, Application US/08788584  
; Patent No. 5837493  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Petithory, Joanne R.  
; TITLE OF INVENTION: NOVEL HUMAN GALECTINS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; FILING DATE: Filed Herewith  
; APPLICATION NUMBER: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0192 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 149 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-788-584-3

Query Match 40.8%; Score 783; DB 2; Length 149;  
Best Local Similarity 97.3%; Pred. No. 5.1e-75;

Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 207 MFSTPAIPPMYHPAYPMPFITTILGGLYPSKILLSGTVLPQAORFHNLCSGNHIAF 266  
Db 1 MFSTXGIPPMYHPGYPMPFITTILGGLYPSKILLSGTVLPQAORFHNLCSGNHIAF 60  
QY 267 HLNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWLCEAHCLKVAVDGQHL 326  
Db 61 HLNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWLCEAHCLKVAVDGQHL 120  
QY 327 FEYHRLNRLPTINRLEVGGDIQLTHVQT 355  
Db 121 FEYHRLNRLPTINRLEVGGDIQLTHVQT 149

## RESULT 3

US-08-788-584-1  
; Sequence 1, Application US/08788584  
; Patent No. 5837493  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Petithory, Joanne R.  
; TITLE OF INVENTION: NOVEL HUMAN GALECTINS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/788,584  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0192 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 145 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-788-584-1

Query Match 34.5%; Score 661; DB 2; Length 145;  
Best Local Similarity 90.5%; Pred. No. 3.6e-62;  
Matches 124; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 219 PHPAYPMPFITTILGGLYPSKILLSGTVLPQAORFHNLCSGNHIAFHLPNPRFDENAV 278  
Db 9 PYLSPXVPFSGTXQGLYPSKILLSGTVLPQAORFHNLCSGNHIAFHLPNPRFDENAV 68  
QY 279 RNTQIDNSWGSEERSLPRKMPFVRGQSFVWLCEAHCLKVAVDGQHLFEYHRLNRLPT 338  
Db 69 RNNQIDNSWGSEERSLPRKMPFVRGQSFVWLCEAHCLKVAVDGQHLFEYHRLNRLPT 128

QY 246 TVLP<sup>SAQR</sup>FHINLC<sup>SGN</sup>- -HIAFHLNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRC 30

Db 213 YVPTGKSFANFKVGSSCDIALHINPRMNGTVVRNSLLNGSWGSEKKITHN-PFGPG 271



Db 272 OPFDLSIRGLDRFYKANGOHLEFDFAHRLSAFORVDTLEIQDVTLSYVQ 322

## RESULT 8

US-08-788-584-5

; Sequence 5, Application US/08788584

; Patent No. 5837493

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Goli, Surya K.

; APPLICANT: Bandman, Olga

; APPLICANT: Hawkins, Phillip R.

; APPLICANT: Pettithory, Joanne R.

; TITLE OF INVENTION: NOVEL HUMAN GALECTINS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/788,584

; FILING DATE: Filed Herewith

; CLASSIFICATION: 436

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0192 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 145 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 727176

US-08-788-584-5

Query Match 27.9%; Score 535; DB 2; Length 145;

Best Local Similarity 70.3%; Pred. No. 7.2e-49;

Matches 104; Conservative 9; Mismatches 29; Indels 6; Gaps 1;

QY 208 FSTPAIPPMYPPHAYPMFPFITILGGLYPSKILLSGTVLPSAQRHINLCSGNHIAFH 267

Db 4 FSTQT-----PYPNLAVPFTSIPNGLYPSKISIVISGVVLSDAKRFQINLCRGGDIAFH 57

QY 268 LNPREDENAVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGOHLF 327

Db 58 LNPREDENAVRNTQINNSWGPEERSLPGSMFPSRGQSFVWILCEGHCFKVAVDGOHC 117

QY 328 EYHRLNLPNTINRLEVGGDIQLTHVQT 355

Db 118 EYSHRLMNLDPDINTLEVAGDIQLTHVET 145

## RESULT 9

US-08-946-914-12

; Sequence 12, Application US/08946914  
; Patent No. 6027916  
; GENERAL INFORMATION:  
; APPLICANT: Ni, Jian  
; APPLICANT: Gentz, Reiner L.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
; STREET: 1100 New York Ave., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/946,914  
; FILING DATE: Herewith  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/028,093  
; FILING DATE: 09-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 145 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
US-08-946-914-12

Query Match 27.9%; Score 535; DB 3; Length 145;

Best Local Similarity 70.3%; Pred. No. 7.2e-49;

Matches 104; Conservative 9; Mismatches 29; Indels 6; Gaps 1;

QY 208 FSTPAIPPMYPPHAYPMFPFITILGGLYPSKILLSGTVLPSAQRHINLCSGNHIAFH 267

Db 4 FSTQT-----PYPNLAVPFTSIPNGLYPSKISIVISGVVLSDAKRFQINLCRGGDIAFH 57

QY 268 LNPREDENAVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGOHLF 327

Db 58 LNPREDENAVRNTQINNSWGPEERSLPGSMFPSRGQSFVWILCEGHCFKVAVDGOHC 117

QY 328 EYHRLNLPNTINRLEVGGDIQLTHVQT 355

Db 118 EYSHRLMNLDPDINTLEVAGDIQLTHVET 145

## RESULT 10

US-09-131-648-5

; Sequence 5, Application US/09131648

; Patent No. 6168920

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Yue, Henry

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl J.

; APPLICANT: Patterson, Chandra

; TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS

; FILE REFERENCE: PF-0576 US

```

; ; REGISTRATION NUMBER: 36,688
; ; REFERENCE/DOCKET NUMBER: 1498.0560001/EKS/SGW
; ; TELECOMMUNICATION INFORMATION:
; ; TELEPHONE: 202-371-2600
; ; TELEFAX: 202-371-2540
; ; INFORMATION FOR SEQ ID NO: 6:
; ; SEQUENCE CHARACTERISTICS:
; ; LENGTH: 317 amino acids
; ; TYPE: amino acid
; ; TOPOLOGY: linear
; ; MOLECULE TYPE: protein
US-08-946-914-6

```

Query Match	24.6%	Score	471.5;	DB	3;	Length	317;
Best Local Similarity	33.0%;	Pred.	No. 1.2e-41;				
Matches	115;	Conservative	55;	Mismatches	128;	Indels	51;
Gaps	8;						
QY	12	SPAVPFGSTGGQLDGLQTIVNGTIVLSSSGTRFVAFNFTQSGFN---	DIAFHENPREFD	68			
Db	14	NPIVIPFGTTIPDQLDPTLLIVIRGHV--PSADRFQVDLVONGSSVKPRADVAFHNPREKR	72				
QY	69	GGYVVCWTRONGSWGPEERKTHMPFOKGMPFDLCFLCYQSDSFVKVMVNGILSEVQFYHVPFF	128				
Db	73	AGCIVCNTLINEKWGREIITYDPFKREKSFEIIMVLKDKFAQVANGKHTLLGXGRIGP	132				
QY	129	HRYVDTTSVNGSVOLSYISFONRTPVPOPAFSTVPVFQSOPVCFPPRRGRRROKPCGVWPAN	188				
Db	133	EKIIDLTYIGYKVNIHSIGF-----SFSDDLQST-----	-QAS	163			
QY	189	PAPITOTVIHTVSAPQGMSTPAIPPMMYPHPAYPMPEITITLGLGYPKSILLSGTVL	248				
Db	164	SLEATEIIVRENVPKS-----GTPL-----	SUPFAARLNTMGPGRTVVVKGEVN	208			
QY	249	PSAQRHFHINICSG--NHIAFHLNPRFDENAVVRVTQIDNSGBERSLUPRKMPFVGRCOSF	306				
Db	209	ANAKSFNVDLLAGSKDIALHLNRLNIKAFVRNSFLQESWGEEERNI-TAFPSPGMYF	267				
QY	307	SWMLCEAHCLKVAVDOGHLEFYHYHRLNRNLPTINLRLEVGGDIQUTHVQT	355				
Db	268	EMIIYCDVREFKVAVNGVSHLEYKHREKEISSIDTLEINGDHILLEVRFS	316				

```

RESULT 12
US-08-728-521-3
; Sequence 3, Application US/08728521
; Patent No. 5869289
;
; GENERAL INFORMATION:
;
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: HUMAN GALECTIN HOMOLOG
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
;
; SOFTWARE: FastSeq Version 1.5
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,521
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749

```

REFERENCE/DOCKET NUMBER: PF-0137 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 316 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 717032  
US-08-728-521-3

Query Match 23.7%; Score 454.5; DB 2; Length 316;  
Best Local Similarity 31.7%; Pred. No. 7.3e-40;  
Matches 114; Conservative 62; Mismatches 133; Indels 51; Gaps 8;

QY 1 MAFSGSQAPLSPAPVPSGTTGGGLQDGLQITVNGTVLSSSGTRFVAVNFOTGFS---GND 57  
DB 2 LSLSLQNIINPTIPIYVSTITEQLKPGSLIVIRGHVPKDS-ERFQVDFQHGNSLKPRAD 60  
QY 58 IAFHENPRFEDGGYVVCNTRQNGSMGPPEKTHMPFQKGMPEFLCFLVQSSDFKVMVNGI 117  
DB 61 VAFHENPREKRSNCIVCNTLTNEKGWEEITHDMPEKSEFIVIMVLKKNFHVAVNGK 120  
QY 118 LFVQYFHRVPRHVDITSVNGSVQLSYISFQNPRTVPVQPAFSTVPFSPVCFPPRPRGR 177  
DB 121 HLLVAHRINPEKIDTLGIFGKVNHSIGR-----FSSDLQSM----- 159  
QY 178 RQKPGVWPANPAPITQVIHTVQSAQPMFSTPAIPPMYPHPAYPPFITTLGGLYP 237  
DB 160 -----ETSLGLTQISKENIQKS-GKLHLS-----LPFEARLNASMG 196  
QY 238 SKSILLSTGLVPSAQRHINLCSG--NHIAFLNPRFDENAVVRNTQIDNSWGSEERSLP 295  
DB 197 GRTVVVKGEVNTNATSFNVDLVAGRSRDIALHLNPLNKAFAVRNSFLQDANGEERNI- 255  
QY 296 RKMFPVQSGFSWILCEAHCLKVAVDGOHLFEYHRLNRLPTINRLEVGDIQTHVOT 355  
DB 256 TCFPFSSGMFEMIIYCDVREFKAVNGVHSLEYKHKFKDLSSIDTLAVDGDGIRLLDVS 315

RESULT 13  
US-08-647-960-2  
Sequence 2, Application US/08647960  
Patent No. 5908761  
GENERAL INFORMATION:  
APPLICANT: ZICK, Yehiel  
TITLE OF INVENTION: GALECTIN-8 AND GALECTIN-8-LIKE PROTEINS  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/647,960  
FILING DATE: 30-MAY-1996  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 107880  
FILING DATE: 05-DEC-1993

ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: ZICK-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3527  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 316 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-647-960-2

Query Match 23.7%; Score 454.5; DB 2; Length 316;  
Best Local Similarity 31.7%; Pred. No. 7.3e-40;  
Matches 114; Conservative 62; Mismatches 133; Indels 51; Gaps 8;

QY 1 MAFSGSQAPLSPAPVPSGTTGGGLQDGLQITVNGTVLSSSGTRFVAVNFOTGFS---GND 57  
DB 2 LSLSLQNIINPTIPIYVSTITEQLKPGSLIVIRGHVPKDS-ERFQVDFQHGNSLKPRAD 60  
QY 58 IAFHENPRFEDGGYVVCNTRQNGSMGPPEKTHMPFQKGMPEFLCFLVQSSDFKVMVNGI 117  
DB 61 VAFHENPREKRSNCIVCNTLTNEKGWEEITHDMPEKSEFIVIMVLKKNFHVAVNGK 120  
QY 118 LFVQYFHRVPRHVDITSVNGSVQLSYISFQNPRTVPVQPAFSTVPFSPVCFPPRPRGR 177  
DB 121 HLLVAHRINPEKIDTLGIFGKVNHSIGR-----FSSDLQSM----- 159  
QY 178 RQKPGVWPANPAPITQVIHTVQSAQPMFSTPAIPPMYPHPAYPPFITTLGGLYP 237  
DB 160 -----ETSLGLTQISKENIQKS-GKLHLS-----LPFEARLNASMG 196  
QY 238 SKSILLSTGLVPSAQRHINLCSG--NHIAFLNPRFDENAVVRNTQIDNSWGSEERSLP 295  
DB 197 GRTVVVKGEVNTNATSFNVDLVAGRSRDIALHLNPLNKAFAVRNSFLQDANGEERNI- 255  
QY 296 RKMFPVQSGFSWILCEAHCLKVAVDGOHLFEYHRLNRLPTINRLEVGDIQTHVOT 355  
DB 256 TCFPFSSGMFEMIIYCDVREFKAVNGVHSLEYKHKFKDLSSIDTLAVDGDGIRLLDVS 315

RESULT 14  
US-08-946-914-15  
Sequence 15, Application US/08946914  
Patent No. 6027916  
GENERAL INFORMATION:  
APPLICANT: Ni, Jian  
APPLICANT: Gentz, Reiner L.  
APPLICANT: Ruben, Steven M.  
TITLE OF INVENTION: Galectin 8, 9, 10 and 10S9  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.  
STREET: 1100 New York Ave., Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3934  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/946,914  
FILING DATE: Herewith  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,093





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2001, 09:42:18 ; Search time 36.57 Seconds  
(without alignments)  
332.532 Million cell updates/sec

Title: US-09-485-951-2

Perfect score: 1917

Sequence: 1 MAFSGSQAPYLPSPVFSCT.....LPTINRLEVGGDIQLTHVQT 355

Scoring table:

Gapop 10.0 , Gapext 0.5

Archived: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1917	100.0	355	LEG9_HUMAN	O00182 homo sapien
2	1374.5	71.7	354	LEG9_RAT	P97840 rattus norv
3	1324	69.1	353	LEG9_MOUSE	O08573 mus musculu
4	603.5	31.5	324	LEG4_RAT	P38552 rattus norv
5	574	29.9	323	LEG4_HUMAN	P56470 homo sapien
6	541	28.2	323	LEG4_PIG	Q29058 sus scrofa
7	535	27.9	144	LEG5_RAT	P47967 rattus norv
8	518	27.0	301	LEG6_MOUSE	O34891 mus musculu
9	467.5	24.4	316	LEG6_MOUSE	Q91115 mus musculu
10	463.5	24.2	316	LEG8_HUMAN	O00214 homo sapien
11	454.5	23.7	316	LEG8_RAT	Q62665 rattus norv
12	427	22.3	283	LEG1_HAECO	O44126 haemochus
13	414.5	21.6	279	LEG3_CABEL	P36573 caenorhabdi
14	374	19.5	285	LEG3_CABEL	Q09581 caenorhabdi
15	339	17.7	244	LEG3_CRITLO	P47953 cricetulus
16	330.5	17.2	295	LEG3_CANFA	P38486 canis famil
17	324.5	16.9	241	LEG3_RABIT	P47845 oryctolagus
18	322.5	16.8	263	LEG3_MOUSE	P16110 mus musculu
19	321	16.7	261	LEG3_RAT	P08699 rattus norv
20	318.5	16.0	245	LEG3_HUMAN	P17931 homo sapien
21	230.5	12.0	135	LEG7_HUMAN	P47929 homo sapien
22	223.5	11.7	135	LEG7_MOUSE	O54974 mus musculu
23	216.5	11.3	135	LEG7_RAT	P97590 rattus norv
24	189.5	9.9	134	LEG6_CHICK	P32668 gallus gall
25	184.5	9.6	134	LEG4_CHICK	P07580 gallus gall
26	182	9.5	182	LEG7_CABEL	Q09605 caenorhabdi
27	167	8.7	134	LEG1_MOUSE	P16045 mus musculu
28	166	8.7	134	LEG1_SHEEP	P81184 ovis aries
29	164	8.6	134	LEG1_CRIGR	P48538 cricetulus
30	164	8.6	134	LEG1_RAT	P11762 rattus norv
31	164	8.6	184	LEG8_CABEL	Q09610 caenorhabdi
32	160	8.3	134	LEG1_BOVIN	P11116 bos taurus
33	155	8.1	129	LEG_ELEEL	P08520 electrophor

RESULT 1

LEG9_HUMAN	154	8.0	134	1	LEG1_HUMAN	P09382 homo sapien
AC	000182; O14532; O75028;				LEG2_CONNY	Q9YIC2 conger myri
DT	01-NOV-1997 (Rel. 35, Created)				LEG2_HUMAN	Q05315 homo sapien
DT	15-JUL-1998 (Rel. 36, Last sequence update)				LEG2_RAT	Q92144 rattus norv
DT	01-OCT-2000 (Rel. 40, Last annotation update)				LEG2_MOUSE	P11946 mus musculu
DE	GALECTIN-9 (HOM-HD-21) (ECALECTIN).				LEG2_HUMAN	P05162 homo sapien
GN	LGALS9.				LEG2_PIG	P36217 bufo arenar
OS	Homo sapiens (Human).				LEG2_PIG	Q29373 sus scrofa
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				LEG2_PIG	Q09277 caenorhabdi
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.				LEG2_PIG	P07645 pseudorabdi
OX	NCBI_TaxID=9606;				LEG2_PIG	P26788 conger myri
[1]					LEG2_PIG	Q09606 caenorhabdi
RN	SEQUENCE FROM N.A. (SHORT FORM).					
RP	TISSUE=Spleen;					
RC	MEDLINE=97197815; PubMed=9045665;					
RA	Tuerci O., Schmitt H., Fadle N., Pfreundschuh M., Sahin U.;					
RT	"Molecular definition of a novel human galectin which is immunogenic					
RT	in patients with Hodgkin's disease.";					
RL	J. Biol. Chem. 272:6416-6422(1997).					
[2]						
RP	SEQUENCE FROM N.A. (LONG FORM).					
RC	TISSUE=Gastric carcinoma;					
RA	Kato S.;					
RL	Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.					
[3]						
RP	SEQUENCE FROM N.A. (SHORT FORM).					
RC	MEDLINE=98307937; PubMed=9642261;					
RA	Matsumoto R., Matsumoto H., Seki M., Asano Y., Kanegasaki S.,					
RA	Stevens R.L., Hirashima M.;					
RT	"Human ealectin, a variant of human galectin-9, is a novel eosinophil					
RT	chemoattractant produced by T lymphocytes.";					
RL	J. Biol. Chem. 273:16976-16984(1998).					
CC	- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-					
CC	EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS. THE					
CC	SHORT ISOFORM ACTS AS AN EOSINOPHIL CHEMOATTRACTANT.					
CC	- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN					
CC	HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.					
CC	- TISSUE SPECIFICITY: PERIPHERAL BLOOD LEUKOCYTES AND LYMPHATIC					
CC	TISSUES. OVEREXPRESSED IN HODGKIN'S DISEASE TISSUE.					
CC	- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING					
CC	DOMAINS.					
CC	- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; Z49107; CAA8922.1; -;					
DR	EMBL; AB006782; BAA22166.1; -;					
DR	EMBL; AB005894; BAA31542.1; -;					

DR HSSP; P17931; 1A3K.  
 DR MIM; 601879; -  
 DR InterPro; IPR001079; -  
 DR Pfam; PF00337; Gal-bind\_lectin; 2.  
 DR PROSITE; PS00309; GALAPTIN; 2.  
 DR Galaptin; Lectin; Repeat; Alternative splicing.  
 KW DOMAIN 1 148 GALAPTIN 1.  
 FT DOMAIN 149 205 LINKER.  
 FT DOMAIN 207 355 GALAPTIN 2.  
 FT BINDING 82 88 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT BINDING 287 293 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT VARSPLIC 149 180 MISSING (IN SHORT ISOFORM).  
 FT CONFLICT 5 5 G -> S (IN REF. 3).  
 FT CONFLICT 88 88 K -> R (IN REF. 1).  
 FT CONFLICT 135 135 S -> F (IN REF. 1).  
 FT CONFLICT 270 270 P -> L (IN REF. 1).  
 FT CONFLICT 313 313 E -> G (IN REF. 1).  
 SQ SEQUENCE 355 AA; 39518 MW; 4748C222FCAFA536A CRC64;

Query Match 100.0%; Score 1917; DB 1; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-146;  
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFSGSQAPYLSPAVPSGTIOGGLQDGLQITVNGTVLSSSGTRFAVNFOTGFGSNDIAF 60  
 DB 1 MAFSGSQAPYLSPAVPSGTIOGGLQDGLQITVNGTVLSSSGTRFAVNFOTGFGSNDIAF 60  
 QY 61 HFNPRDGGYVVCNTRQNSWGPGEKTHMPFKGMPDCLFVQSSDFKVMVNGILFV 120  
 DB 61 HFNPRDGGYVVCNTRQNSWGPGEKTHMPFKGMPDCLFVQSSDFKVMVNGILFV 120  
 QY 121 QYFHRVPFHRVDTSVNGSVQLSYISFQNPRTVPQPAFSTVPFSPQVCFPPRGRROK 180  
 DB 121 QYFHRVPFHRVDTSVNGSVQLSYISFQNPRTVPQPAFSTVPFSPQVCFPPRGRROK 180  
 QY 181 PGVWPANPAPIQTQVIHTVQSPAGQMFSTPAIPPMYPHPAYPMPFITTLGLGLYPSKS 240  
 DB 181 PGVWPANPAPIQTQVIHTVQSPAGQMFSTPAIPPMYPHPAYPMPFITTLGLGLYPSKS 240  
 QY 241 ILLSGTVLPSAQRFHNLCSGNHIAFLNPRFDENAVRNTQIDNSWGSSEERSLPRKMPF 300  
 DB 241 ILLSGTVLPSAQRFHNLCSGNHIAFLNPRFDENAVRNTQIDNSWGSSEERSLPRKMPF 300  
 QY 301 VRQGSFVWLCEAHLCKVAVDQHLFEYHRLNRLPTINRLEVGDDIQLTHVQT 355  
 DB 301 VRQGSFVWLCEAHLCKVAVDQHLFEYHRLNRLPTINRLEVGDDIQLTHVQT 355

SULT 2  
 9\_RAT  
 AC LEG9\_RAT STANDARD; PRT; 354 AA.  
 DT P97840; O08588; O35866;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GALECTIN-9 (36 KDA BETA-GALACTOSIDE BINDING LECTIN) (URATE  
 DE TRANSPORTER/CHANNEL) (UAT).  
 GN LGALS9.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RP [1]  
 RP SEQUENCE FROM N.A. (LONG AND SHORT FORMS).  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Kidney;  
 RX MEDLINE=97150351; PubMed=9038233;  
 RA Wada J., Kanwar Y.S.;  
 RT "Identification and characterization of galectin-9, a novel beta-  
 RT galactoside-binding mammalian lectin.";  
 RL J. Biol. Chem. 272:6078-6086(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A. (SHORT FORM).

RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Kidney;  
 RX MEDLINE=97150769; PubMed=8995305;  
 RA Leal-Pinto E., Tao W., Rappaport J., Richardson M., Knorr B.A.,  
 RA Abramson R.G.;  
 RT "Molecular cloning and functional reconstitution of a urate  
 RT transporter/channel.";  
 RL J. Biol. Chem. 272:617-625(1997).  
 CC -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-  
 CC EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS (BY  
 CC SIMILARITY). MAY PROVIDE THE MOLECULAR BASIS FOR URATE FLUX ACROSS  
 CC CELL MEMBRANES, ALLOWING URATE THAT IS FORMED DURING PURINE  
 CC METABOLISM TO EFFLUX FROM CELLS AND SERVING AS AN ELECTROGENIC  
 CC TRANSPORTER THAT PLAYS AN IMPORTANT ROLE IN RENAL AND  
 CC GASTROINTESTINAL URATE EXCRETION. HIGHLY SELECTIVE TO THE ANION  
 CC URATE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-  
 CC CLASSICAL SECRETORY PATHWAY (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN  
 CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE  
 CC SMALL INTESTINE.  
 CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING  
 CC DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; U59462; AAB51192.1; -  
 CC EMBL; U72741; AAB68592.1; -  
 CC EMBL; U67958; AAB48591.1; -  
 CC HSSP; P17931; 1A3K.  
 DR InterPro; IPR001079; -  
 DR Pfam; PF00337; Gal-bind\_lectin; 2.  
 DR PROSITE; PS00309; GALAPTIN; 2.  
 KW Galaptin; Lectin; Repeat; Alternative splicing; Ion transport.  
 FT DOMAIN 1 147 GALAPTIN 1.  
 FT DOMAIN 148 205 LINKER.  
 FT DOMAIN 206 354 GALAPTIN 2.  
 FT BINDING 81 87 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT BINDING 286 292 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT VARSPLIC 148 179 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 354 AA; 39946 MW; 6574F960B2EAF37C CRC64;

Query Match 71.7%; Score 1374.5; DB 1; Length 354;  
 Best Local Similarity 71.8%; Pred. No. 3.7e-103;  
 Matches 255; Conservative 31; Mismatches 68; Indels 1; Gaps 1;

QY 1 MAFSGSQAPYLSPAVPSGTIOGGLQDGLQITVNGTVLSSSGTRFAVNFOTGFGSNDIAF 60  
 DB 1 MAFSGSQAPYLSPAVPSGTIOGGLQDGLQITVNGTVLSSSGTRFAVNFOTGFGSNDIAF 60  
 QY 61 HFNPRDGGYVVCNTRQNSWGPGEKTHMPFKGMPDCLFVQSSDFKVMVNGILFV 120  
 DB 61 HFNPRDGGYVVCNTRQNSWGPGEKTHMPFKGMPDCLFVQSSDFKVMVNGILFV 120  
 QY 121 QYFHRVPFHRVDTSVNGSVQLSYISFQNPRTVPQPAFSTVPFSPQVCFPPRGRROK 180  
 DB 121 QYFHRVPFHRVDTSVNGSVQLSYISFQNPRTVPQPAFSTVPFSPQVCFPPRGRROK 180  
 QY 181 PGVWPANPAPIQTQVIHTVQSPAGQMFSTPAIPPMYPHPAYPMPFITTLGLGLYPSKS 240  
 DB 181 PGVWPANPAPIQTQVIHTVQSPAGQMFSTPAIPPMYPHPAYPMPFITTLGLGLYPSKS 240  
 QY 241 ILLSGTVLPSAQRFHNLCSGNHIAFLNPRFDENAVRNTQIDNSWGSSEERSLPRKMPF 300  
 DB 241 ILLSGTVLPSAQRFHNLCSGNHIAFLNPRFDENAVRNTQIDNSWGSSEERSLPRKMPF 300  
 QY 301 VRQGSFVWLCEAHLCKVAVDQHLFEYHRLNRLPTINRLEVGDDIQLTHVQT 355  
 DB 301 VRQGSFVWLCEAHLCKVAVDQHLFEYHRLNRLPTINRLEVGDDIQLTHVQT 355

```

QY 301 VRQSFVSWILCEAHCLKVAVDQGHLEFYHYHRLNLPPTINRLEVGGDIQLTHVQT 355
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 300 NRQGSFVSWILCEBCHGCFKAVDQGHICEYHYHRLKNLPDINTLEVAGDIQLTHVQT 354

RESULT 3
LEG9_MOUSE STANDARD; PRT; 353 AA.
ID AC 008573; 008572;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GALECTIN-9.
GN LGALS9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RP STRAIN=CD-1; TISSUE=Small intestine, and Kidney;
RX MEDLINE=97190351; PubMed=9038233;
RA Wada J., Kanwar Y.S.;
RT "Identification and characterization of galectin-9, a novel beta-
RL galactoside-binding mammalian lectin.";
RL J. Biol. Chem. 272:6078-6086(1997).
RN [2]
CHARACTERIZATION.
RX MEDLINE=97298141; PubMed=9153289;
RA Wada J., Ota K., Kumar A., Wallner E.I., Kanwar Y.S.;
RT "Developmental regulation, expression, and apoptotic potential of
RL galectin-9, a beta-galactoside binding lectin.";
RL J. Clin. Invest. 99:2452-2461(1997).
CC -1- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-
CC EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-
CC CLASSICAL SECRETORY PATHWAY.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: ACCENTUATED EXPRESSION IN LIVER AND THYMUS OF
CC EMBRYO, DETECTED IN EMBRYONIC HEART, BRAIN, LUNG, LIVER, AND
CC KIDNEY. HIGHLY EXPRESSED IN ADULT THYMUS, SMALL INTESTINE, AND
CC LIVER, AND TO A LESSER EXTENT IN LUNG, KIDNEY, SPLEEN, CARDIAC,
CC AND SKELETAL MUSCLE. BARELY DETECTABLE IN BRAIN AND RETICULOCYTE.
CC THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE SMALL INTESTINE.
CC -1- DEVELOPMENTAL STAGE: THE EXPRESSION INCREASED WITH SUCCESSIVE
CC STAGES OF EMBRYONIC DEVELOPMENT.
CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U55061; AAB51190.1; -
CC EMBL; U55060; AAB51189.1; -
CC DR HSSP; P17931; 1A3K.
CC DR MGD; MGI:109496; Lgals9.
CC InterPro: IPR001079; -
CC DR Pfam; PF00337; Gal_bind_lectin; 2.
CC DR PROSITE; PS00309; GALAPTIN; 2.
CC DR Galaptin; Lectin; Repeat; Alternative splicing.
CC FT DOMAIN 1 147 GALAPTIN 1.
CC FT DOMAIN 148 204 LINKER.
CC FT DOMAIN 205 353 GALAPTIN 2.
CC FT BINDING 81 87
CC FT BINDING 285 291
CC FT BINDING 285 291 BETA-GALACTOSIDE (BY SIMILARITY).
CC FT BINDING 285 291 BETA-GALACTOSIDE (BY SIMILARITY).

```

```
CC -----
DR EMBL; M73553; AAA41505.1; -.
DR PIR; A46631; A46631.
DR HSSP; PI7931; IA3K.
DR InterPro; IPR001079; -.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
KW Galaptin; Lectin; Repeat.
FT DOMAIN 1 152
FT DOMAIN 153 177
FT DOMAIN 178 324
FT BINDING 257 263
FT BINDING 257 263
SQ SEQUENCE 324 AA; 36347 MW; 478024D7322AFE7B CRC64;

Query Match 31.5%; Score 603.5; DB 1; Length 324;
Best Local Similarity 39.6%; Pred. No. 2.le-41;
Matches 139; Conservative 54; Mismatches 115; Indels 43; Gaps 10;

QY 9 PYLSPAVPFGTIOGDLQGLQITVNGTVLSSGTRFAVNFQTG-FSGNDIAFHFNPRE 67
11 PTYNPLPKRPIPGSLVGMSSIIQG-TAKDNMRHFVNFAVGQDEGADIAFHFNPRE 69
68 DGGYVVCNTRQSGWGPPEERKTHMPFQKMPDCLFVQSSDFKVMVNGILFVQYFHRVP 127
70 GWDKVVNTMQSGWCKEKKKSMPEKQGHFELVEMVHSEHYKVVVNGTPEYEGHRLP 129
128 FHRVDTSVNGSVQLSYISFQNPRTVPVQPAFSTVPFQPCFPVPPRRGRKPPGWPA 187
130 LQWVTHLQVGDGLQSLINF-----LGGQFAASQYPTMTI-----PA 167
188 NPAPITQTVIHTVQSAPOGFSTPAI--PPMYPHPAYPMPFTITLGLYPSKILLSG 245
168 YP-----SAGYPPQMSLVPWAGPIFNP-----PPIVGTIQQGLTARRTIIGK 214
246 TVLPSAQRHINLCGN--HIAFLNPRFENAVVNTQIDNSWGSEERSLPRKMPFVRG 303
215 YVLPATAKLIINFKVSGTGDIAPHMNPRIQD-CVVRNSYMWNGSWGSEERKIPYN-PFGAG 272
304 QSFVSWILCEAHLKVAVDQHLFEYFHYHRLNLPINRLEVGDDIOLTHVQ 354
273 QPFDLIRCGTDRFKVYANGQHLFDFSHRFAQFQVRVDMLEIKGDTLSYVQ 323

RESULT 5
LEG4_HUMAN
ID LEG4_HUMAN STANDARD; PRT; 323 AA.
AC P56470;
DT 15-JUL-1998 (Rel. 36, Created)
15-JUL-1998 (Rel. 36, Last sequence update)
15-JUL-1998 (Rel. 36, Last annotation update)
DE GALECTIN-4 (LACTOSE-BINDING LECTIN 4) (L-36 LACTOSE BINDING PROTEIN)
DE (L36LBP).
GN LGALS4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal carcinoma;
RA Kato S.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal carcinoma;
RA Rechreche H., Mallo G.V., Montalto G., Dagorn J.C., Iovanna J.L.;
RT "Cloning and expression of the mRNA of human galectin-4, an S-type
RT lectin down-regulated in colorectal cancer.";
RL Eur. J. Biochem. 248:225-230(1997).
CC -!- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
CC SUGARS. MAY BE INVOLVED IN THE ASSEMBLY OF ADHERENS JUNCTIONS.
```

```
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB006781; BAA22165.1; -.
DR EMBL; AF014838; AAC51763.1; -.
DR HSSP; PI7931; IA3K.
DR MIM; 602518; -.
DR InterPro; IPR001079; -.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
KW Galaptin; Lectin; Repeat.
FT DOMAIN 1 152
FT DOMAIN 153 177
FT DOMAIN 178 323
FT BINDING 256 262
FT BINDING 256 262
SQ SEQUENCE 323 AA; 35941 MW; E79BC0A9AB3990EF CRC64;

Query Match 29.9%; Score 574; DB 1; Length 323;
Best Local Similarity 38.2%; Pred. No. 4.8e-39;
Matches 134; Conservative 52; Mismatches 121; Indels 44; Gaps 10;

QY 9 PYLSPAVPFGTIOGDLQGLQITVNGTVLSSGTRFAVNFQTGFS-GNDIAFHFNPRE 67
11 PTYNPLPKRPIPGSLVGMSSIIQG-VASEHMKRFFVNVVQDGPDSVAFHFNPRE 69
68 DGGYVVCNTRQSGWGPPEERKTHMPFQKMPDCLFVQSSDFKVMVNGILFVQYFHRVP 127
70 GWDKVVNTMQSGWCKEKKKSMPEKQGHFELVEMVHSEHYKVVVNGTPEYEGHRLP 129
128 FHRVDTSVNGSVQLSYISFQNPRTVPVQPAFSTVPFQPCFPVPPRRGRKPPGWPA 187
130 LQWVTHLQVGDGLQSLINFQGG-----PLRPGQ-----PPMMP 165
188 NPAPITQTVIHTVQSAPOGFSTPAI--PPMYPHPAYPMPFTITLGLYPSKILLSG 245
166 YFQP-----GHCHQ-----QLNSLPTMEGPTFNP-----PVPYFGRLLQGLTARRTIIGK 212
246 TVLPSAQRHIN--LCSGNHIAFLNPRFENAVVNTQIDNSWGSEERSLPRKMPFVRG 303
213 YVPTGKSAINFKVSGSGDIALHINPRMGNGTVVRNSLNGSWGSEERKITHN-PFGPG 271
304 QSFVSWILCEAHLKVAVDQHLFEYFHYHRLNLPINRLEVGDDIOLTHVQ 354
272 QPFDLIRCGTDRFKVYANGQHLFDFHRLSAFQVRVDMLEIQGVDTLSYVQ 322

RESULT 6
LEG4_PIG
ID LEG4_PIG STANDARD; PRT; 323 AA.
AC Q29058;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE GALECTIN-4 (LACTOSE-BINDING LECTIN 4) (L-36 LACTOSE BINDING PROTEIN)
DE (L36LBP).
GN LGALS4.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
```

RC TISSUE-Tongue;  
 RX MEDLINE-95081129; PubMed-7989350;  
 RA Chiu M.L., Parry D.A.D., Feldman S.R., Klapper D.G., O'Keefe E.J.;  
 RT "An adherens junction protein is a member of the family of lactose-  
 binding lectins";  
 RL J. Biol. Chem. 269:31770-31776(1994).  
 RN [2]  
 RP SEQUENCE OF 1-140 FROM N.A.  
 RC TISSUE-Small intestine;  
 RX MEDLINE-96327607; PubMed-9672129;  
 RA Winteroe A.K., Fredholm M., Davies W.;  
 RT "Evaluation and characterization of a porcine small intestine cDNA  
 library: analysis of 839 clones";  
 RL Mamm. Genome 7:509-517(1996).  
 CC -1- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF  
 CC SUGARS. MAY BE INVOLVED IN THE ASSEMBLY OF ADHERENS JUNCTIONS.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING  
 CC DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X79303; CAA53884.1; -;  
 DR EMBL; F14653; CAA23179.1; -;  
 DR HSSP; P17931; 1A3K.  
 DR InterPro: IPR001079; -;  
 DR Pfam: PF00337; Gal-bind lectin; 2.  
 DR PROSITE; PS00309; GALAPTIN; FALSE\_NEG.  
 KW Galaptin; Lectin; Repeat.  
 FT DOMAIN 1 152 GALAPTIN 1.  
 FT DOMAIN 153 177 LINKER.  
 FT DOMAIN 178 323 GALAPTIN 2.  
 FT BINDING 256 262 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT BINDING 98 99 AP -> GA (IN REF. 2).  
 FT CONFLICT 126 126 H -> T (IN REF. 2).  
 FT CONFLICT 126 126 H -> T (IN REF. 2).  
 SQ SEQUENCE 323 AA; 35852 MW; 728C761712B29343 CRC64;  
 -----  
 Query Match 28.28; Score 541; DB 1; Length 323;  
 Best Local Similarity 36.55; Pred. No. 2.le-36;  
 Matches 132; Conservative 54; Mismatches 128; Indels 48; Gaps 11;  
 Qy 1 MAF---SGSQAPYLSPAVPFSGTGGQDGLQITVNGTVLSSGTRFAVNFQTGFS-GN 56  
 Db 1 MAFVPAPGYQTY-NPTLPYKYPFGGLRVGMSVYIQG-VANEHMKRFVNVVQGPGEA 58  
 Qy 57 DIAFHENRFRDGGYVVCNTRQNSWGPPEERKTHMPFQKMPFDLCFLVQSDFKVMVNG 116  
 Db 59 DVAFENRFRDGGYVVCNTRQNSWGPPEERKTHMPFQKMPFDLCFLVQSDFKVMVNG 118  
 Qy 117 ILFVQYHRVFRHVRDTSVNGSVQLSVISFQNPRTVPVQAFSTVPSPQVPPRPG 176  
 Db 119 DPFEYFGRIRIVQLVTHLQVDGDLTLQSFNF-----IGQGAPAS----- 157  
 Qy 177 RQKPPGVPVNPAPITQTVIHTVQSAPQGFSTPAIPPMVYPHAY--PMPFTITLGG 234  
 Db 158 -----PGMP-NPG-----YPGCKNQPCNLPCEGAPTNPVYPYKTRLOGG 201  
 Qy 235 LYPKSKILLSGTVLPSAQRFIN--LCSGNHIAFLNPRFDENAVRNTQIDNSGSEER 292  
 Db 202 LVARRTIVIKGVPPSGKSLVINFKVSGSDVALHINPRLEGTIVRNSYLNKGWAEER 261  
 Qy 293 SLPRKMPVRCQGSFVWILCEAHLKAVDQHLFEYFHRLENLPTINRLEVGGDIQTH 352  
 Db 262 KSFN-PPAPGQYFDLSIRCLDRKFKVYANGQHLFDHSHRLSNFQGVDTLETQGDVTL 320

Qy 353 VQ 354  
 Db 321 VQ 322  
 RESULT 7  
 ID LEG5\_RAT STANDARD; PRT; 144 AA.  
 AC P47967;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE GALECTIN-5 (RL-18).  
 GN LGALS5.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE-Reticulocytes;  
 RX MEDLINE-95197487; PubMed-7890611;  
 RA Gitt M.A., Wiser M.F., Leffler H., Herrmann J., Xia Y.-R.,  
 RA Massa S.M., Mapping D.N.W., Lusis A.J., Barondes S.H.;  
 RT "Sequence and mapping of galectin-5, a beta-galactoside-binding  
 RT lectin, found in rat erythrocytes";  
 RL J. Biol. Chem. 270:5032-5038(1995).  
 CC -1- FUNCTION: MAY FUNCTION IN ERYTHROCYTE DIFFERENTIATION.  
 CC -1- SUBUNIT: MONOMER.  
 CC -1- TISSUE SPECIFICITY: ERYTHROCYTES.  
 CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; L36862; AAC42050.1; -;  
 DR EMBL; L21711; AAA65445.1; -;  
 DR HSSP; P17931; 1A3K.  
 DR InterPro: IPR001079; -;  
 DR Pfam: PF00337; Gal-bind lectin; 1.  
 DR PROSITE; PS00309; GALAPTIN; 1.  
 KW Galaptin; Lectin; Acetylation.  
 FT INIT\_MET 0 0  
 FT MOD\_RES 1 1 ACETYLATION.  
 FT BINDING 76 82 BETA-GALACTOSIDE (POTENTIAL).  
 FT CONFLICT 127 135 MISSING (IN AAA65445).  
 SQ SEQUENCE 144 AA; 16065 MW; BC95283D760DA515 CRC64;  
 -----  
 Query Match 27.9%; Score 535; DB 1; Length 144;  
 Best Local Similarity 70.3%; Pred. No. 2.5e-36;  
 Matches 104; Conservative 9; Mismatches 29; Indels 6; Gaps 1;  
 Qy 208 FSTPAIPPMVPHPAYPMPFTITLGGLYPSKILLSGTVLPSAQRFINLCSGNHIAFH 267  
 Db 3 FSTQT-----PYPNLAVPFTTIPNGLYPSKIVISGVLSDAKRQFQINLRCGGDIAFH 56  
 Qy 268 LNPRFDENAVRNTQIDNSGSEERSLPRKMPFVRCQSFVWILCEAHLKAVDQHLF 327  
 Db 57 LNPRFDENAVRNTQIDNSGSEERSLPRKMPFVRCQSFVWILCEAHLKAVDQHLF 327  
 Qy 328 EYFHRLENLPTINRLEVGGDIQTHVQET 355  
 Db 117 EYFHRLENLPTINRLEVGGDIQTHVQET 144  
 RESULT 8  
 LEG6\_MOUSE

ID LEG6\_MOUSE STANDARD; PRT; 301 AA.  
 AC O54891; O88352;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GALECTIN-6.  
 GN LGALS6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=98112847; PubMed=9446608;  
 RA Gitt M.A., Colnot C., Poirier F., Nani K.J., Barondes S.H.,  
 RA Leffler H.;  
 RT "Galectin-4 and galectin-6 are two closely related lectins expressed  
 RT in mouse gastrointestinal tract.";  
 RL J. Biol. Chem. 273:2954-2960(1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=98112848; PubMed=9446609;  
 RA Gitt M.A., Xia Y.-R., Atchison R.E., Lusis A.J., Barondes S.H.,  
 RA Leffler H.;  
 RT "Sequence, structure, and chromosomal mapping of the mouse Lgals6  
 RT gene, encoding galectin-6.";  
 RL J. Biol. Chem. 273:2961-2970(1998).  
 CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING  
 CC DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF026799; AAC04508.1;  
 DR EMBL; AF026796; AAC04508.1; JOINED.  
 DR EMBL; AF026797; AAC04508.1; JOINED.  
 DR EMBL; AF026798; AAC04508.1; JOINED.  
 DR EMBL; AF026794; AAC27244.1;  
 DR MGD; MGI:107335; Lgals6.  
 DR InterPro; IPR001079;  
 DR Pfam; PF00337; Gal-bind\_lectin; 2.  
 DR Galaptin; Lectin; Repeat.  
 DR DOMAIN 1 151 GALAPTIN 1.  
 FT DOMAIN 152 160 LINKER.  
 FT DOMAIN 160 301 GALAPTIN 2.  
 FT CONFLICT 154 154 A -> V (IN AAC27244).  
 SQ SEQUENCE 301 AA; 34112 MW; 9A4DD09944EDFAB9 CRC64;

Query Match 27.0%; Score 518; DB 1; Length 301;  
 Best Local Similarity 35.6%; Pred. No. 1.4e-34;  
 Matches 124; Conservative 52; Mismatches 110; Indels 62; Gaps 10;

QY 9 PYLSPAVPFGSTIQGLQDLOITVNGTVLSSSGTRFAVNFOTG-FSGNDIAFHFNPRE 67  
 DB 11 PTYNPLPKRPITPGSLVGMSEFYIGTA-KENMRFRHFNFAVGDDGADVAHFHNPRED 69  
 QY 68 DGVYVNCNTRQMSWGPEERKTHMPFQKMPFDLCFLVQSSDFKVMVNGILFVQYFHRYP 127  
 DB 70 GWDKVVNTKQSGRWCKEEKS-MPFQKGHFELVPMVPEHKYKVVVNGSPFVEYGHRLP 128  
 QY 128 HRVDITSVNGSVOLSYISFQNPRTVPVQPAFSTVPFSPVCPFPPRGRROKPPGWPA 187  
 DB 129 VQMVTHLQVGDGLQSLINF-----FGVQPAETKYP----- 159

QY 188 NPAPITQTVIHTVQAPGQMFSTPAIPPMYPMPFPIITILGGLYPSKILLSGTV 247  
 DB 160 -----AMTG-----PPVFN-----CLPYVVALOGGFTVRTTIKGV 193  
 QY 248 LPSAORFHIN--LCSGNHIAFHLNPRFDENAVVNTQIDNSWGSERSLPRKMPFVRGOS 305  
 DB 194 LPTAKTFAINFRVGSSEIDIALHINPRIGD-CLVRNSYMNWSGTEERMVAYN-PFGPGQF 251  
 QY 306 FSVWILCEAHCLKVAVDGOHLEFYVHRLNRLPTINRLEVGGDIOITHV 353  
 DB 252 FDLSTRCGMDRFKVFANGIHLEFNFSHRFOALRKINTLEINGDITUSIV 299  
 RESULT 9  
 LEG8\_MOUSE  
 ID LEG8\_MOUSE STANDARD; PRT; 316 AA.  
 AC Q9JUL15;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GALECTIN-8 (LGALS-8).  
 GN LGALS8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL6/J; TISSUE=Kidney;  
 RA Maier C., Haeussler J., Roesch K., Moschagath E., Haeussler J.,  
 RA Vogel W.;  
 RT "The human Lgals-8 gene: genomic sequence and expression of the  
 RT prostate carcinoma tumour antigen (PCTA-1) and the Po66 carbohydrate  
 RT binding protein.";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: POSSESSES SUGAR BINDING AND HEMAGGLUTINATION ACTIVITY.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF218069; AAF27645.1;  
 DR MGD; MGI:1928481; Lgals8.  
 DR InterPro; IPR001079;  
 DR PROSITE; PS00309; GALAPTIN; 1.  
 DR Galaptin; Lectin; Repeat.  
 DR DOMAIN 1 153 GALAPTIN 1.  
 FT DOMAIN 154 184 LINKER.  
 FT DOMAIN 185 316 GALAPTIN 2.  
 FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).  
 SQ SEQUENCE 316 AA; 36161 MW; 11A20309AEF52C69 CRC64;

Query Match 24.4%; Score 467.5; DB 1; Length 316;  
 Best Local Similarity 32.4%; Pred. No. 1.6e-30;  
 Matches 113; Conservative 61; Mismatches 124; Indels 51; Gaps 8;

QY 12 SPAPVPSGTIQGLQDLOITVNGTVLSSSGTRFAVNFOTGFS---GNDIAFHFNPRE 68  
 DB 13 NPILPYVGITITQLKPGSLIVIRGHVPKDS-ERFQVDFQLGNSLKPRAVAFHFNPRFR 71  
 QY 69 GGVVVCNTRQMSWGPEERKTHMPFQKMPFDLCFLVQSSDFKVMVNGILFVQYFHRVPE 128  
 DB 72 SSCIVCNTLTQKKGWEEITYDMPFERKESFEIVFVNLKFKQAVNGRHVLLIYHRISP 131  
 QY 129 HRVDITSVNGSVOLSYISFQNPRTVPVQPAFSTVPFSPVCPFPPRGRROKPPGWPA 188  
 DB 132 EQIDTVGIYKVINHSIGFR-----FSSDLQSM-----ETS 162

QY 189 PAPITQTVIHTVQSAPOQMFSTPAIPPMYHPAYPMPTITILGGLYPSKSIILLSGFVL 248  
 DB 163 ALGLTQINRENIQK-PGKL-----QLSLPFEARLNASMGPGRTVVIKGEVN 207  
 QY 249 PSAORFHINLCSG--NHIAFHLNPRFDENAVVNTQIDNSWGSSEERSLPRKMPFVRGQSF 306  
 DB 208 TNARSFNDLVAGTRDIALHLNRLNKAFRNSFQDANGEEERNI-TCFPFSSGMYF 266  
 QY 307 SWILCEAHCLKVAVDGQHLFEYHRLNRLPTINRLEVGDDIQLTHVOT 355  
 DB 267 EMIIYCDVREKVAINGVHSLEYKHKRFDLSSIDTSLSDGDIRLLDVR 315  
 RESULT 10  
 LEG8\_RAT  
 ID LEG8\_HUMAN STANDARD: PRT: 316 AA.  
 000214: 015215; Q9UP34; Q9UE26; Q9UP33; Q9UP32;  
 01-NOV-1997 (Rel. 35, Created)  
 01-OCT-2000 (Rel. 40, Last sequence update)  
 DE GALECTIN-8 (GAL-8) (PROSTATE CARCINOMA TUMOR ANTIGEN) (PCTA-1)  
 DE (PO66 CARBOHYDRATE-BINDING PROTEIN) (PO66-CBP).  
 GN LGALS8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Prostate;  
 RX MEDLINE=96293510; PubMed=8692978;  
 RA Su Z.-Z., Lin J., Shen R., Fisher P.E., Goldstein N.I., Fisher P.B.;  
 RT "Surface-epitope masking and expression cloning identifies the human  
 RT prostate carcinoma tumor antigen gene PCTA-1 a member of the galectin  
 RT gene family.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 93:7252-7257(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Hippocampus;  
 RA Hadari Y.R., Eisenstein M., Zakut R., Zick Y.;  
 RT "Galectin-8: on the road from structure to function.";   
 RL Trends Glycosci. Glycotechnol. 9:103-112(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE-Lung carcinoma;  
 RA Brichory F., Biron N., Desrues B., Bourguet P., Le Pennec J.P.,  
 RT "Dazord L.;  
 RT "Molecular cloning of a beta-galactoside-binding lectin related to  
 RT galectin-8 and identified in human lung carcinoma.";   
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Maier C., Haessler J., Roesch K., Moschagath E., Vogel W.;  
 RT "Genomic organization and expression of the human galectin-8 gene.";   
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).  
 CC -1- TISSUE SPECIFICITY: SELECTIVE EXPRESSION BY PROSTATE CARCINOMAS  
 CC VERSUS NORMAL PROSTATE AND BENIGN PROSTATIC HYPERTROPHY.  
 CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING  
 CC DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L78132; AAB51605.1; -;  
 DR EMBL; X91790; CAA62904.1; ALT\_INIT.

DR EMBL; AF074000; AAD45402.1; -;  
 DR EMBL; AF074001; AAD45403.1; -;  
 DR EMBL; AF074002; AAD45404.1; -;  
 DR EMBL; AF193806; AAF19370.1; ALT\_INIT.  
 DR EMBL; AF193805; AAF19370.1; JOINED.  
 DR HSSP; P17931; IAA3K.  
 DR InterPro; IPR001079; -;  
 DR PROSITE; PS00337; Gal-bind lectin; 2.  
 DR Galaptin; Lectin; Repeat; Antigen; Alternative splicing; Polymorphism.  
 FT DOMAIN 1 153 GALAPTIN 1.  
 FT LINKER.  
 FT DOMAIN 154 184 GALAPTIN 2.  
 FT DOMAIN 185 316 GALAPTIN 2.  
 FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT VARSPLIC 182 182 L -> LPSNRGGDISKIAPRTVYTKSKDSTVNHHTLTCTKIP  
 FT VARSPLIC 182 182 PMNVSK (IN ISOFORM 2).  
 FT VARSPLIC 182 182 L -> LQTVSPSWLQGGHSETFCFCSVLWTRVFLFIACRPI  
 FT VARSPLIC 182 182 GLTVASFO (IN ISOFORM 3).  
 FT VARIANT 35 35 /FTID=VAR\_009710.  
 FT CONFLICT 18 18 Y -> F (IN REF. 1 AND 2).  
 FT CONFLICT 55 55 V -> M (IN REF. 2).  
 FT CONFLICT 97 99 KRE -> OKER (IN REF. 2).  
 FT CONFLICT 111 111 D -> A (IN REF. 2).  
 FT CONFLICT 170 170 S -> V (IN REF. 1).  
 FT CONFLICT 183 183 S -> R (IN REF. 2).  
 FT CONFLICT 203 203 K -> Q (IN REF. 1).  
 SQ SEQUENCE 316 AA; 35538 MW; 1724C36085A0510A CRC64;  
 Query Match 24.2%; Score 463.5; DB 1; Length 316;  
 Best Local Similarity 31.8%; Pred. No. 3.3e-30;  
 Matches 112; Conservative 52; Mismatches 131; Indels 57; Gaps 7;  
 QY 12 SPAPFSGTIOGGLODGLQITVNGTVLSSSTRAFNAPVQTGFSGN---DIAFHNPRED 68  
 DB 13 NPVIPPYGTIPDQLDPGTLIVICGHV-PSDADRFQDLQNGSSVVKPRADVAFHENPREKR 71  
 QY 69 GGVVVCNTRONGSWGPERKTHMPGKMPDFLCFLQSSDFKVMVNGILFVQFHRVPPF 128  
 DB 72 AGCIVCNTLINEKGREEITYDTFKEKSEFIVIMVLKDKFQVAVNGKHTLLYGHRIGP 131  
 QY 129 HRVDTISVNGSVQLSYISFQNPRTVPVQPAFTVPFSPQVCFPPRPRGRROKPGCVMPAN 188  
 DB 132 EKIDTLGIYGVNIHSIGF-----SPSS----- 154  
 QY 189 PAPITQTVIHTVQSAPOQMFSTPAIPPMYHP---PAYPMPTITILGGLYPSKSIILLSG 245  
 DB 155 -----DLQSTQASSLELTSIRENVKSGTLPQLSPFAARLNTPMGRTVVVVKG 204  
 QY 246 TVLPFAORFHINLCSG--NHIAFHLNPRFDENAVVNTQIDNSWGSSEERSLPRKMPFVRG 303  
 DB 205 EVNNAKSFNVDLVAGSKSDIALHLNRLNKAFRNSFQDANGEEERNI-TSFPSPG 263  
 QY 304 QSFVSWILCEAHCLKVAVDGQHLFEYHRLNRLPTINRLEVGDDIQLTHVOT 355  
 DB 264 MYFEMIYCDVREKVAINGVHSLEYKHKRFDLSSIDTSLSDGDIRLLDVR 315  
 RESULT 11  
 LEG8\_RAT  
 ID LEG8\_RAT STANDARD: PRT: 316 AA.  
 AC Q62665;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE GALECTIN-8 (30 KDA S-TYPE LECTIN) (RL-30).  
 GN LGALS8.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=101116;  
 RN [1]





```

132 DRISVNGSVOLSYISRQNPRTVPVQAFSTVPFSDPVCPFPFRGRQRKPPGGWVPANPAP 191
   :::| | ::| | ::| |
128 SHLSIDGDLYLNHVH-----WGKG--- 146

192 ITQTVIHTVSAGPMGFSPAIPPMYPHPAYDMPFITTTILGGLYPSKSTLLSGTLPSPA 251
   ||:||: | ::| | ::| | ::| | ::| | ::| |
147 -----YYPVPFSGLANGPLVGKSULVFETVEKKA 176

252 QRFHNLCSGN-HIAFHNLPRFDENAVVRNTQIDNSWGSEERSLPRKMPEVROGSFVWI 310
   :||:| | | ::| | ||||| :||:| | ||||| :||:| |
177 KRFHVNLRKRNGDISPFENRDEKHVIRNSLAANEWNREER--GNPFKEGVGDFDLVI 234

311 LCEAHCLKVAVDGQHUFYYHYRLNRPLPTINRLEVGDDIQLTHVQ 354
   | : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
235 QNEEYAFQVFVANGERYISFAHR-ADPHDIAGLIQISGDIELSGIQ 277

RESULT 14
LE33_CAEEEL STANDARD; PRT; 285 AA.
AC Q09581;
DT 01-OCT-1996 (Rel. 34, Created)
DR DT 15-DEC-1998 (Rel. 37, Last sequence update)
DD DT 15-DEC-1998 (Rel. 37, Last annotation update)
DB DE 32 KDA BETA-GALACTOSIDE-BINDING LECTIN LEC-3 (32 KDA GBP).
    LEC-3 OR ZK892.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RR STRAIN=BRISTOL N2;
RA Lloyd C.;
RC Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
RLL [2]
RR REVISIONS.
RR STRAIN=BRISTOL N2;
RA Jones S.J.M.;
RR Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RLL -! FUNCTION: BINDS GALACTOSE (BY SIMILARITY).
CCC -! SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z48638; CAAB88570.1; -.
DDR HSP; P17931; IAK3.
DDR WornPeep; ZK892.1; CEI18468.
DDR InterPro; IPR001079; -.
DDR Pfam; PF00337; Gal-bind_lectin; 2.
DDR PROSITE; PS00309; GALAPTIN; 1.
KKW Galaptin; Lectin; Repeat; Multigene family.
FT DOMAIN 1 137 GALAPTIN 1.
FF DOMAIN 138 285 GALAPTIN 2.
FF BINDING 212 218 BETA-GALACTOSIDE (BY SIMILARITY).
SSO SEQUENCE 285 AA; 32465 MW; 5D191988B760BAAE CRC64;

Query Match          19.5%; Score 374; DB 1; Length 285;
Best Local Similarity 28.1%; Pred. No. 4.4e-23;
Matches 97; Conservative 54; Mismatches 112; Indels 82; Gaps 10;

QY 15 VPSGTIOGLODGLOITVNGTVTLSSGTRFYAWE--QTGFGNDIAFHNPREFDGGY 71
DB 9 IPYSKLTERIEPGQTLTKTGKTIDES-KRFNLIHKHSDFSGNDVPLHLSTRF-DEGK 66
Y 72 VVCNTRONSWGPEEKTHMFKOGKMFCDLCFLVSSDKFMVNGILFVOYFHRVPPHRV 131

```

Db 67 IYVNAVTKGTWKEERAKN-PIKKGDDDIRAHSDSKFQVSIHKEVNFHRIPLNSV 125  
QY 132 DTISVNGSVOLSYISQNPRTVPQAFSTVPFSQVCPFRPRGRQRKPPGWPANPAP 191  
Db 126 SHLSIDGDVVLNHHVQ-----WGCK-----144  
QY 192 ITQTVIHTVQSAPGOMFSTPAIPPMYPHPAYPMFITTILG-GLYPSKILLSGTVLPS 250  
Db 145 -----YYPVYESGIAADGLVPGKTLVVYGTPEKK 174  
QY 251 AQRFHINLCSGN-HIAFHLNPRFDENAVRNTQIDNSWGESEERSLPKMPFVGRGSFSVM 309  
Db 175 AKKFENLKKNGKIDALHENPRFDEKSVVRNSLVNGEWNEERE--GKNPFERLTAFDLE 232  
QY 310 ILCEAHCLVAVDGOHLFEYVYHRLNLPINRLEVGDDIQLTHVQ 354  
Db 233 IRNEEFAQIFVNGERFASYAHRV-DPHDIAGLIQGDIELTGIQ 276

RESULT 15

CRILLO STANDARD; PRT; 244 AA.  
AC P47953;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE GALECTIN-3 (GALACTOSE-SPECIFIC LECTIN 3) (MAC-2 ANTIGEN) (IGE-BINDING  
DE PROTEIN) (35 KDA LECTIN) (CARBOHYDRATE BINDING PROTEIN 35) (CBP 35)  
DE (LAMININ-BINDING PROTEIN) (LECTIN L-29) (CBP30).  
GN LGALS3.  
OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetus.  
OX NCBI\_TaxID=10030;  
RN  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RX MEDLINE=94299546; PubMed=8027086;  
RA Mehul B., Bawumia S., Martin S.R., Hughes R.C.;  
RT "Structure of baby hamster kidney carbohydrate-binding protein CBP30,  
RT an S-type animal lectin.";  
RL J. Biol. Chem. 269:18250-18258(1994).  
CC -!- FUNCTION: GALACTOSE-SPECIFIC LECTIN WHICH BINDS IGF.  
CC -!- SIMILARITY: THE C-TERMINAL DOMAIN BELONGS TO THE GALAPTIN  
CC (S-LECTIN) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X78879; CAA55479.1; .  
DR HSP: P17931; I43K.  
DR InterPro: IPR001079; .  
DR Pfam: PF00337; Gal-bind\_lectin; 1.  
DR PROSITE: PS00309; GALAPTIN; 1.  
KW Galaptin; Lectin; IGF-binding protein; Repeat; Phosphorylation;  
KW Acetylation.  
FT INIT\_MET 0 0 BY SIMILARITY.  
FT MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
FT MOD\_RES 5 5 PHOSPHORYLATION (BY CK1)  
FT FT (BY SIMILARITY).  
FT DOMAIN 34 98 7 X 9 AA TANDEM REPEATS OF Y-P-G-X(3)-P-  
FT [GS]-A.  
FT REPEAT 34 42 1.  
FT REPEAT 43 51 2.  
FT REPEAT 52 60 3.  
FT REPEAT 61 69 4.

FT REPEAT 70 77 5 (APPROXIMATE).  
FT REPEAT 78 87 6 (APPROXIMATE).  
FT REPEAT 88 98 7 (APPROXIMATE).  
FT DOMAIN 112 244 GALAPTIN.  
FT DISULFID 167 167 INTERCHAIN (BY SIMILARITY).  
FT BINDING 175 181 BETA-GALACTOSIDE (BY SIMILARITY).  
SQ SEQUENCE 244 AA; 25608 MW; 8F99B9AA0BBA7D3F CRC64;  
Query Match 17.7%; Score 339; DB 1; Length 244;  
Best Local Similarity 38.9%; Pred. No. 2.3e-20;  
Matches 88; Conservative 26; Mismatches 84; Indels 28; Gaps 9;  
QY 149 NRTVP-----VQPAESTVP-FSQPVCFPPR-PRGR--RQKPPGVPA-----NPAPI 192  
Db 17 NQGNPFGWGNQPGAGGYPCASYPGAYPCQAPPQAYPGQAPGAYPCGTAPGAYPCGAPG 76  
QY 193 TQT-----VIHTVQSAPGOMFSTPAIPPMYPHPAYPMFITTILGSLYPSKILLSGTVL 248  
Db 77 AYPGQPGASGAYPSAPG---AYPAAGPYCAPTGALTVPYKPLAGGVMPRLITTINGTVK 133  
QY 249 PSAQRFHINLCSGNHIAFHLNPRFDEN---AVVRNTQIDNSWGESEERSLPKMPFVRGQS 305  
Db 134 PNANRIILNFLRGNDAIEHFNPRFNNRRVIVCNTKQDNNGREERQ--SAPFESGRP 191  
QY 306 FSVWILCEAHCLKLVAVDGOHLFEYVYHRLNLPINRLEVGDDIQLT 351  
Db 192 FKIQVLVEADHFKVAVNDHALLIQYHNRKMLNREINOMEISGDIILT 237

Search completed: August 1, 2001, 09:42:19  
Job time: 684 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2001, 09:30:55 ; Search time 36.57 Seconds  
(without alignments)  
29.975 Million cell updates/sec

Title: US-09-485-951-1  
Perfect score: 178  
Sequence: 1 NPRTVPVQPAFTVPFSPQVCPFRPRGRQK 32

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Archived: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	178	100.0	355	1 LEG9_HUMAN	O00182 homo sapien
2	104	58.4	354	1 LEG9_RAT	P97840 rattus norv
3	94	52.8	353	1 LEG9_MOUSE	O08573 mus musculu
4	62.5	35.1	806	1 MK07_MOUSE	Q9wvs8 mus musculu
5	55	30.9	1490	1 CRK7_HUMAN	Q9yvv4 homo sapien
6	54	30.3	487	1 EBN2_EBV	P12978 epstein-bar
7	53.5	30.1	360	1 NO44_SOYBN	P04672 glycine max
8	52.5	29.5	791	1 ARNT_MOUSE	P53762 mus musculu
9	51.5	28.9	169	1 HSB7_MOUSE	P35385 mus musculu
10	51.5	28.9	970	1 PSU1_YEAST	P53550 saccharomyc
11	51	28.7	460	1 MYCN_MARMO	Q61976 saccharom
12	51	28.7	1172	1 TSP2_HUMAN	P35442 homo sapien
13	51	28.7	1429	1 L112_CABEL	P14585 caenorhabdi
14	50.5	28.4	90	1 HSB7_RAT	Q9uk5 rattus norv
15	50.5	28.4	489	1 CPCML_RAT	P19225 rattus norv
16	50.5	28.4	800	1 ARNT_RAT	P41739 rattus norv
17	50	28.1	1139	1 RBL2_HUMAN	O08999 homo sapien
18	50	28.1	1400	1 RON_HUMAN	O04912 homo sapien
19	49.5	27.8	324	1 GC1_MOUSE	P01868 mus musculu
20	49.5	27.8	393	1 GC1M_MOUSE	P01869 mus musculu
21	49.5	27.8	464	1 MYCN_HUMAN	P04198 homo sapien
22	49.5	27.8	551	1 IL2B_HUMAN	P14784 homo sapien
23	49.5	27.8	872	1 S145_HUMAN	Q13435 homo sapien
24	49	27.5	417	1 HEM2_SELMA	P45623 selaginella
25	49	27.5	434	1 CPCO_RAT	P33273 rattus norv
26	49	27.5	1135	1 RBL2_MOUSE	Q64700 mus musculu
27	49	27.5	1135	1 RBL2_RAT	O55081 rattus norv
28	49	27.5	1175	1 DSRA_RAT	P55266 rattus norv
29	49	27.5	1906	1 KMLS_CHICK	P11799 gallus gall
30	48.5	27.2	170	1 HSB7_HUMAN	Q9ubv9 homo sapien
31	48.5	27.2	324	1 MATD_NEUCR	Q10116 neurospora
32	48.5	27.2	524	1 FTWH_MYCTU	O06223 mycobacteri
33	48.5	27.2	790	1 ARNT_RABIT	O02748 oryctolagus

RESULT 1

ID	LEG9_HUMAN	STANDARD;	PRT;	355 AA.
AC	O00182; O14532; O75028;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	GALECTIN-9 (HOM-HD-21) (ECALECTIN).			
GN	LGALS9.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (SHORT FORM).			
RC	TISSUE=Splice;			
RX	MEDLINE=97197815; PubMed=9045665;			
RA	Tuerceci O., Schmitt H., Fadle N., Pfreundschuh M., Sahin U.;			
RT	"Molecular definition of a novel human galectin which is immunogenic			
RT	in patients with Hodgkin's disease.";			
RL	J. Biol. Chem. 273:16976-16984(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A. (LONG FORM).			
RC	TISSUE=Gastric carcinoma;			
RA	Kato S.;			
RL	Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A. (SHORT FORM).			
RX	MEDLINE=98307937; PubMed=9642261;			
RA	Matsumoto R., Matsumoto H., Seki M., Asano Y., Kanegasaki S.,			
RA	Stevens R.L., Hirashima M.;			
RT	"Human ecalectin, a variant of human galectin-9, is a novel eosinophil			
RT	chemoattractant produced by T lymphocytes.";			
RL	J. Biol. Chem. 273:16976-16984(1998).			
CC	- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-			
CC	EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS. THE			
CC	SHORT ISOFORM ACTS AS AN EOSINOPHIL CHEMOATTRACTANT.			
CC	- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN			
CC	HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.			
CC	- TISSUE SPECIFICITY: PERIPHERAL BLOOD LEUKOCYTES AND LYMPHATIC			
CC	TISSUES. OVEREXPRESSED IN HODGKIN'S DISEASE TISSUE.			
CC	- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING			
CC	DOMAINS.			
CC	- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; 249107; CAA88922.1;			
DR	EMBL; AB006782; BAA22166.1;			
DR	EMBL; AB005894; BAA31542.1;			

34	48	27.0	152	1	HMGB_SOYBN	Q10370 glycine max
35	48	27.0	153	1	VPG_BYDVP	P09513 barley yell
36	48	27.0	313	1	GDA7_WHEAT	P04727 triticum ae
37	48	27.0	379	1	PSPB_DICDI	P54704 dictyosteli
38	48	27.0	391	1	AZAB_ERIEU	O15012 erinaceus e
39	48	27.0	409	1	SERA_ECOLI	P08328 escherichia
40	48	27.0	450	1	AZAB_HUMAN	P18089 homo sapien
41	48	27.0	476	1	TR16_HUMAN	Q15654 homo sapien
42	48	27.0	559	1	3BP2_MOUSE	Q05649 mus musculu
43	48	27.0	564	1	PKNA_ANASP	P54734 anabaena sp
44	48	27.0	1044	1	BUBL_SCHPO	O94751 schizosacch
45	47.5	26.7	176	1	HMGA_SOYBN	Q00423 glycine max

ALIGNMENTS

DR HSSP; PI7931; 1A3K.  
 DR MIM; 601879; -  
 DR InterPro; IPR001079; -  
 DR Pfam; PF00337; Gal-bind-lectin; 2.  
 DR PROSITE; PS00309; GALAPTIN; 2.  
 KW Galaptin; Lectin; Repeat; Alternative splicing.  
 FT DOMAIN 1 148 GALAPTIN 1.  
 FT DOMAIN 149 206 LINKER.  
 FT DOMAIN 207 355 GALAPTIN 2.  
 FT BINDING 82 88 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT BINDING 287 293 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT VARSPPLIC 149 180 MISSING (IN SHORT ISOFORM).  
 FT CONFLICT 5 5 G -> S (IN REF. 3).  
 FT CONFLICT 88 88 K -> R (IN REF. 1).  
 FT CONFLICT 135 135 S -> F (IN REF. 1).  
 FT CONFLICT 270 270 P -> L (IN REF. 1).  
 FT CONFLICT 313 313 E -> G (IN REF. 1).  
 SQ SEQUENCE 355 AA; 4748C22FCAFA536A CRC64;

Query Match 100.0%; Score 178; DB 1; Length 355;  
 Best Local Similarity 100.0%; Pred. No. 4.8e-15;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVPVQPAFSTVPFSQPVCFPPRRGRQK 32  
 |||||  
 Db 149 NPTVPVQPAFSTVPFSQPVCFPPRRGRQK 180

## RESULT 2

LEG9\_MOUSE  
 ID LEG9\_MOUSE STANDARD; PRT; 354 AA.  
 AC P97840; O08588; O35866;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GALECTIN-9 (36 KDA BETA-GALACTOSIDE BINDING LECTIN) (URATE  
 TRANSPORTER/CHANNEL) (UAT).  
 GN LGALS9.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (LONG AND SHORT FORMS).  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney, and Small intestine;  
 RX MEDLINE=97190351; PubMed=9038233;  
 RA Wada J., Kanwar Y.S.;  
 RT "Identification and characterization of galectin-9, a novel beta-  
 galactoside-binding mammalian lectin.";  
 RL J. Biol. Chem. 272:6078-6086(1997).  
 [2]

SEQUENCE FROM N.A. (SHORT FORM).  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;  
 RX MEDLINE=97150769; PubMed=8995305;  
 RA Leal-Pinto E., Tao W., Rappaport J., Richardson M., Knorr B.A.,  
 RA Abramson R.G.;  
 RT "Molecular cloning and functional reconstitution of a urate  
 transporter/channel.";  
 RL J. Biol. Chem. 272:617-625(1997).  
 CC -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-  
 EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS (BY  
 SIMILARITY). MAY PROVIDE THE MOLECULAR BASIS FOR URATE FLUX ACROSS  
 CELL MEMBRANES, ALLOWING URATE THAT IS FORMED DURING PURINE  
 METABOLISM TO EFFLUX FROM CELLS AND SERVING AS AN ELECTROGENIC  
 TRANSPORTER THAT PLAYS AN IMPORTANT ROLE IN RENAL AND  
 GASTROINTESTINAL URATE EXCRETION. HIGHLY SELECTIVE TO THE ANION  
 URATE.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-  
 CLASSICAL SECRETORY PATHWAY (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN  
 HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -!- TISSUE SPECIFICITY: THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE

CC SMALL INTESTINE.  
 CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING  
 CC DOMAINS.  
 CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; U59462; AAB51192.1; -  
 DR EMBL; U72741; AAB68592.1; -  
 DR EMBL; U67958; AAB48591.1; -  
 DR HSSP; PI7931; 1A3K.  
 DR InterPro; IPR001079; -  
 DR Pfam; PF00337; Gal-bind-lectin; 2.  
 DR PROSITE; PS00309; GALAPTIN; 2.  
 KW Galaptin; Lectin; Repeat; Alternative splicing; Ion transport.  
 FT DOMAIN 1 147 GALAPTIN 1.  
 FT DOMAIN 148 205 LINKER.  
 FT DOMAIN 206 354 GALAPTIN 2.  
 FT BINDING 81 87 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT BINDING 286 292 BETA-GALACTOSIDE (BY SIMILARITY).  
 FT VARSPPLIC 148 179 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 354 AA; 39946 MW; 6574F960B2EAF37C CRC64;

Query Match 58.4%; Score 104; DB 1; Length 354;  
 Best Local Similarity 59.4%; Pred. No. 5.8e-06;  
 Matches 19; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 NPTVPVQPAFSTVPFSQPVCFPPRRGRQK 32  
 | |||||  
 Db 148 NSTAAPQVPVFSTMQFSQPVQFPPMPKGRQK 179

## RESULT 3

LEG9\_MOUSE  
 ID LEG9\_MOUSE STANDARD; PRT; 353 AA.  
 AC O08573; O08572;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE GALECTIN-9.  
 GN LGALS9.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Small intestine, and Kidney;  
 RX MEDLINE=97190351; PubMed=9038233;  
 RA Wada J., Kanwar Y.S.;  
 RT "Identification and characterization of galectin-9, a novel beta-  
 galactoside-binding mammalian lectin.";  
 RL J. Biol. Chem. 272:6078-6086(1997).  
 [2]  
 RN CHARACTERIZATION.  
 RP MEDLINE=97298141; PubMed=9153289;  
 RX Wada J., Ota K., Kumar A., Wallner E.I., Kanwar Y.S.;  
 RA "Developmental regulation, expression, and apoptotic potential of  
 galectin-9, a beta-galactoside binding lectin.";  
 RL J. Clin. Invest. 99:2452-2461(1997).  
 CC -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-  
 EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-  
 CLASSICAL SECRETORY PATHWAY.  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN  
 HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.





[illegible]





use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X53673; CAA37712.1; -  
EMBL; X53674; CAA37712.1; JOINED.

DR InterPro: IPR001092; -

DR InterPro: IPR002418; -

DR InterPro: IPR003015; -

DR Pfam: PF00010; HLH; 1.

DR Pfam: PF01056; MYC\_N-term; 1.

DR PRINTS: PR00044; LEUZIPPRMYC.

DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; 1.

KW Nuclear protein; DNA-binding; Proto-oncogene; Phosphorylation.

FT DOMAIN 258 274 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 377 390 BASIC DOMAIN.

FT DOMAIN 391 430 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).

FT DOMAIN 429 450 LEUCINE-ZIPPER (POTENTIAL).

FT MOD\_RES 257 257 PHOSPHORYLATION (BY CK2)

FT MOD\_RES 259 259 (BY SIMILARITY).

FT MOD\_RES 259 259 PHOSPHORYLATION (BY CK2)

FT SEQUENCE 460 AA; 49192 MW; 8A16686C82F5B02E CRC64;

Query Match 28.7%; Score 51; DB 1; Length 460;

Best Local Similarity 38.5%; Pred. No. 24;

Matches 15; Conservative 1; Mismatches 13; Indels 10; Gaps 1;

QY 1 NPRTVPQPAFS-----TVPFSQVCPFPFRGR 29

Db 200 DPAPVPVAPAGSPAVGAAGAAAPASAAVAPRLLGR 238

RESULT 12

TSP2\_HUMAN

ID TSP2\_HUMAN STANDARD; PRT; 1172 AA.

AC P35442;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE THROMBOSPONDIN 2 PRECURSOR.

GN THBS2 OR TSP2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94010892; PubMed=8406456;

RA Labell T.L., Byers P.H.;

RT "Sequence and characterization of the complete human thrombospondin 2

cDNA: potential regulatory role for the 3' untranslated region."

RL Genomics 17:225-229(1993).

RN [2]

RP SEQUENCE OF 560-1172 FROM N.A.

RC TISSUE=Fibroblast;

RX MEDLINE=92217961; PubMed=1559694;

RA Labell T.L., McGookey Milewicz D.J., Distche C.M., Byers P.H.;

RT "Thrombospondin II: partial cDNA sequence, chromosome location, and

expression of a second member of the thrombospondin gene family in

humans."

RL Genomics 12:421-429(1993).

CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN.

CC -!- SUBUNIT: HOMOTRIMER. CROSS-LINKED BY DISULFIDE BONDS.

CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.

CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.

CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.

CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L12350; AAA03703.1; -

EMBL; M81339; -; NOT\_ANNOTATED\_CDS.

PIR; A42173; A42173.

DR HSP; P00740; LIIXA.

DR MIM; 188061; -

DR InterPro: IPR000561; -

DR InterPro: IPR000884; -

DR InterPro: IPR001007; -

DR Pfam: PF00008; EGF; 2.

DR Pfam: PF00090; tsp\_1; 3.

DR Pfam: PF00093; vwc; 1.

DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.

DR PROSITE: PS01186; EGF\_2; 1.

DR PROSITE: PS50092; TSP1; 3.

DR PROSITE: PS01208; VWFC; 1.

KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;

KW EGF-like domain; Signal.

FT SIGNAL 1 18

FT CHAIN 19 1172

FT DOMAIN 19 232

FT DOMAIN 318 375

FT DOMAIN 381 432

FT DOMAIN 437 493

FT DOMAIN 494 548

FT DOMAIN 549 589

FT DOMAIN 590 647

FT DOMAIN 648 692

FT DOMAIN 725 760

FT DOMAIN 761 783

FT DOMAIN 784 819

FT DOMAIN 820 842

FT DOMAIN 843 880

FT DOMAIN 881 916

FT DOMAIN 917 952

FT DOMAIN 953 1172

FT SITE 928 930

FT SITE 266 266

FT DISULFID 270 270

FT DISULFID 553 564

FT DISULFID 558 574

FT DISULFID 577 588

FT DISULFID 594 610

FT DISULFID 601 619

FT DISULFID 622 646

FT DISULFID 652 665

FT DISULFID 659 678

FT DISULFID 680 691

FT CARBOHYD 151 151

FT CARBOHYD 316 316

FT CARBOHYD 330 330

FT CARBOHYD 457 457

FT CARBOHYD 584 584

FT CARBOHYD 710 710

FT CARBOHYD 1089 1089

FT SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;

SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;

Query Match 28.7%; Score 51; DB 1; Length 1172;

Best Local Similarity 44.8%; Pred. No. 60;

Matches 13; Conservative 2; Mismatches 8; Indels 6; Gaps 2;

QY 2 PRTVPQPAFSTVPFSQVCPFPFRGR 30

Db 608 PRCVNTQPGFHCLP-----C-PPRYRGNQ 630

RESULT 13  
 ID L112 CAEBL STANDARD; PRT; 1429 AA.  
 AC P14595;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE LIN-12 PROTEIN PRECURSOR.  
 GN LIN-12 OR R107.8  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=88334747; PubMed=3419531;  
 RA Yochem J., Weston K., Greenwald I.;  
 RT "The Caenorhabditis elegans lin-12 gene encodes a transmembrane  
 protein with overall similarity to Drosophila Notch.";  
 Nature 335:547-550(1988).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RC MEDLINE=94150718; PubMed=7906398;  
 RA Willson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
 RA Sims M., Smaldon N., Smith A., Smith M., Sonnenhammer E., Staden K.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans";  
 Nature 368:32-38(1994).  
 RL  
 CC -1- FUNCTION: LIN-12 IS INVOLVED IN SEVERAL CELL FATES DECISIONS THAT  
 CC REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12  
 CC ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES  
 CC EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: HIGH, TO C.ELEGANS GLP-1.  
 CC -1- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.  
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; M12069; AAA70191.1; -;  
 CC EMBL; Z14092; CAA78474.1; -;  
 CC PIR; S06434; S06434.  
 CC HSSP; P00740; 11XA.  
 CC WormPep; R107.8; CE00274.  
 CC InterPro; IPR000152; -;  
 CC InterPro; IPR000561; -;  
 CC InterPro; IPR000800; -;  
 CC InterPro; IPR001881; -;  
 CC InterPro; IPR002110; -;  
 CC Pfam; PF00008; EGF; 13.  
 CC Pfam; PF00023; ank; 4.  
 CC Pfam; PF00066; notch; 3.  
 CC PROSITE; PS50088; ANK\_REPEAT; 3.  
 -----  
 DR PROSITE; PS50297; ANK\_REPEAT; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 3.  
 DR PROSITE; PS00022; EGF\_1; 12.  
 DR PROSITE; PS01186; EGF\_2; 11.  
 DR PROSITE; PS01187; EGF\_CA; 2.  
 KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;  
 KW Glycoprotein; Signal.  
 FT SIGNAL 1 15  
 FT CHAIN 16 1429  
 FT LIN-12 PROTEIN.  
 FT DOMAIN 16 908  
 FT EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 909 931  
 FT POTENTIAL.  
 FT DOMAIN 932 1429  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 20 61  
 FT EGF-LIKE 1.  
 FT DOMAIN 114 150  
 FT EGF-LIKE 2.  
 FT DOMAIN 152 190  
 FT EGF-LIKE 3.  
 FT DOMAIN 201 246  
 FT EGF-LIKE 4.  
 FT DOMAIN 250 285  
 FT EGF-LIKE 5.  
 FT DOMAIN 287 323  
 FT EGF-LIKE 6.  
 FT DOMAIN 323 363  
 FT EGF-LIKE 7.  
 FT DOMAIN 365 402  
 FT EGF-LIKE 8.  
 FT DOMAIN 404 441  
 FT EGF-LIKE 9.  
 FT DOMAIN 449 492  
 FT EGF-LIKE 10.  
 FT DOMAIN 503 541  
 FT EGF-LIKE 11.  
 FT DOMAIN 543 579  
 FT EGF-LIKE 12.  
 FT DOMAIN 582 619  
 FT EGF-LIKE 13.  
 FT DOMAIN 635 669  
 FT LIN/NOTCH 1.  
 FT REPEAT 670 710  
 FT LIN/NOTCH 2.  
 FT REPEAT 711 750  
 FT LIN/NOTCH 3.  
 FT REPEAT 1093 1122  
 FT ANK 1.  
 FT REPEAT 1126 1158  
 FT ANK 2.  
 FT REPEAT 1162 1194  
 FT ANK 3.  
 FT REPEAT 1206 1236  
 FT ANK 4.  
 FT REPEAT 1240 1269  
 FT ANK 5.  
 FT DISULFID 24 35  
 FT BY SIMILARITY.  
 FT DISULFID 29 49  
 FT BY SIMILARITY.  
 FT DISULFID 51 60  
 FT BY SIMILARITY.  
 FT DISULFID 118 129  
 FT BY SIMILARITY.  
 FT DISULFID 123 138  
 FT BY SIMILARITY.  
 FT DISULFID 140 149  
 FT BY SIMILARITY.  
 FT DISULFID 156 169  
 FT BY SIMILARITY.  
 FT DISULFID 163 178  
 FT BY SIMILARITY.  
 FT DISULFID 180 189  
 FT BY SIMILARITY.  
 FT DISULFID 205 227  
 FT BY SIMILARITY.  
 FT DISULFID 221 234  
 FT BY SIMILARITY.  
 FT DISULFID 236 245  
 FT BY SIMILARITY.  
 FT DISULFID 254 264  
 FT BY SIMILARITY.  
 FT DISULFID 259 273  
 FT BY SIMILARITY.  
 FT DISULFID 275 284  
 FT BY SIMILARITY.  
 FT DISULFID 291 302  
 FT BY SIMILARITY.  
 FT DISULFID 296 311  
 FT BY SIMILARITY.  
 FT DISULFID 313 322  
 FT BY SIMILARITY.  
 FT DISULFID 327 339  
 FT BY SIMILARITY.  
 FT DISULFID 334 351  
 FT BY SIMILARITY.  
 FT DISULFID 353 362  
 FT BY SIMILARITY.  
 FT DISULFID 369 381  
 FT BY SIMILARITY.  
 FT DISULFID 375 390  
 FT BY SIMILARITY.  
 FT DISULFID 392 401  
 FT BY SIMILARITY.  
 FT DISULFID 408 419  
 FT BY SIMILARITY.  
 FT DISULFID 413 429  
 FT BY SIMILARITY.  
 FT DISULFID 431 440  
 FT BY SIMILARITY.  
 FT DISULFID 507 518  
 FT BY SIMILARITY.  
 FT DISULFID 512 529  
 FT BY SIMILARITY.  
 FT DISULFID 531 540  
 FT BY SIMILARITY.  
 FT DISULFID 547 558  
 FT BY SIMILARITY.  
 FT DISULFID 552 567  
 FT BY SIMILARITY.  
 FT DISULFID 569 578  
 FT BY SIMILARITY.  
 FT DISULFID 586 597  
 FT BY SIMILARITY.  
 FT DISULFID 591 607  
 FT BY SIMILARITY.  
 FT DISULFID 609 618  
 FT BY SIMILARITY.  
 FT CARBOHYD 41 41  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 194 194  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 378 378  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1429 AA; 157115 MW; 255EDD7A62C025DB CRC64;

Query Match 28.7%; Score 51; DB 1; Length 1429;

Best Local Similarity 54.5%; Pred. No. 72;

Matches 12; Conservative 0; Mismatches 6; Indels 4; Gaps 1;

QY 12 STVPFSQPVCF-----PPRRGR 29

Db 196 STVEFKQPVCFLEISADHPDGR 217

## RESULT 14

ID HSB7\_RAT STANDARD; PRT; 90 AA.  
AC Q90UK5;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE HEAT-SHOCK PROTEIN, BETA-7 (CARDIOVASCULAR HEAT SHOCK PROTEIN) (CVHSP)  
DE (FRAGMENT).  
GN HSPB7 OR CVHSP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=20062883; PubMed=10593960;  
RA Krief S., Faivre J.-F., Robert P., Le Douarin B., Brument-Larignon N.,  
RA Lefrere I., Bouzyk M.M., Anderson K.M., Greller L.D., Tobin F.L.,  
RA Souchet M., Bril A.;  
RT Identification and characterization of cvhsp. A novel human small  
RT stress protein selectively expressed in cardiovascular and  
RT insulin-sensitive tissues.;  
RL J. Biol. Chem. 274:36592-36600(1999).  
CC -!- SUBUNIT: INTERACTS WITH C-TERMINAL DOMAIN OF ACTIN-BINDING PROTEIN  
CC 280 (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: FOUND IN BOTH CARDIAC AND SKELETAL MUSCLE.  
CC -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)  
CC FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AJ243193; CAB63268.1; -;  
DR EMBL; AF155910; AAF20024.1; -;  
DR PROSITE; PS01031; HSP20; PARTIAL.  
KW Heat shock; Chapterone.  
FT DOMAIN 1 8 POLY-SER.  
FT NON\_TER 90 90  
SQ SEQUENCE 90 AA; 9804 MW; 9D122DB6443ED1FE CRC64;

Query Match 28.4%; Score 50.5; DB 1; Length 90;

Best Local Similarity 33.3%; Pred. No. 5.7;

Matches 11; Conservative 6; Mismatches 7; Indels 9; Gaps 1;

QY 6 PVQPAFS-----TVPFSQPVCFPPRRGR 29

Db 17 PMEKALSMFSEDFGSMPLHSEPLTFPPARPGQ 49

## RESULT 15

ID CPCM\_RAT STANDARD; PRT; 489 AA.  
AC P19225;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE CYTOCHROME P450 2C22 (EC 1.14.14.1) (CYP1C22) (P450 MD) (P450 P49).  
GN CYP2C22 OR CYP2C-22 OR P450MD.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;  
RX MEDLINE=90370497; PubMed=2395662;  
RA Nagata K., Sasamura H., Miyata M., Shimada M., Yamazoe Y., Kato R.;  
RT "cDNA and deduced amino acid sequences of a male dominant P-450md  
RT mRNA in rats";  
RL Nucleic Acids Res. 18:4934-4934(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;  
RX MEDLINE=91088589; PubMed=2263625;  
RA Emi Y., Chijiwa C., Omura T.;  
RT "A different cytochrome P450 form is induced in primary cultures of  
RT rat hepatocytes";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:9746-9750(1990).  
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE  
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN  
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY  
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY  
CC ACIDS, AND XENOBIOTICS.  
CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +  
CC OXIDIZED FLAVOPROTEIN + H(2)O.  
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X53477; CAA37570.1; -;  
DR EMBL; M58041; AAA40950.1; -;  
DR PIR; S11160; S11160.  
DR PIR; A39257; A39257.  
DR InterPro; IPR001128; -;  
DR InterPro; IPR002401; -;  
DR Pfam; PF00067; P450; 1.  
DR PRINTS; PR00385; P450.  
DR PRINTS; PR00463; EP4501.  
DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;  
KW Microsome; Endoplasmic reticulum.  
FT BINDING 434 434 HEME.  
FT CONFLICT 50 50 I -> M (IN REF. 2).  
FT CONFLICT 177 177 V -> L (IN REF. 2).  
FT CONFLICT 456 456 T -> I (IN REF. 2).  
FT CONFLICT 478 478 V -> L (IN REF. 2).  
FT CONFLICT 483 483 E -> Q (IN REF. 2).  
SQ SEQUENCE 489 AA; 56156 MW; E4C5B082A81CB4D CRC64;

Query Match 28.4%; Score 50.5; DB 1; Length 489;

Best Local Similarity 44.4%; Pred. No. 29;

Matches 12; Conservative 2; Mismatches 10; Indels 3; Gaps 1;

Qy 1 NPR--TVPQAFSTVPFQVCFPP 24  
| | | | | : | | |  
Db 462 NPEDIDTTPVQGLLSVPPPELGFIP 488

Search completed: August 1, 2001, 09:42:18  
Job time: 683 sec

OM of: US-09-485-951-1 to: EST:\* out\_format : pfs

Date: Aug 1, 2001 12:16 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q/cg2\_1/USPTO.spool/US09485951/runat\_31072001\_202942\_13211/app\_query.fasta\_1.504  
-DB=EST -QMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=6.000  
-GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELETE=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=ext -MINLEN=0  
-ALIGN=15 -MODE=LOCAL -OUTEXT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=2000000000 -USER=US09485951\_GCGN1\_1\_3162 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLIFY -WAIT -THREADS=1

Search information block:

Query: US-09-485-951-1

Query length: 32

Database: EST:\*

Database sequences: 10228115

Database length: 431459454

Search time (sec): 1962.150000

score\_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
gb_est46:AW403229	+	178.00	370.43	1.6e-11	439	AW403229 UI-HF-BKO-abb-a-03-0-0
gb_est53:AW86539	+	171.00	350.83	2.0e-10	780	AW86539 RC1-OP0083-220300-021
gb_gss9:AQ279515	-	170.00	354.59	1.2e-10	394	AQ279515 C17BI-E1-25111111.TR.C1
gb_est84:AW373817	+	164.00	337.52	1.1e-09	665	AW373817 QV3-BF0537-221299-048
gb_est29:AL559254	+	152.00	314.23	2.2e-08	535	AL559254 AL559254 LTI-NFL008.TC
gb_est39:BF883419	+	118.00	245.36	0.0002	404	BF883419 QV3-ET0203-081200-525
gb_est84:BF143767	+	106.00	215.45	0.0070	703	BF143767 601789554F1.NCI.CGAP.1
gb_est84:BF181157	+	106.00	213.25	0.0032	909	BF181157 601805820F1.NCI.CGAP.M
gb_est89:BF533977	+	106.00	211.12	0.0123	1165	BF533977 602075139F1.NCI.CGAP.1
gb_est82:BF046316	+	104.00	217.93	0.0051	323	BF046316-BP250002BI0C9.Soares
gb_est50:AW55379	+	104.00	213.54	0.0089	539	AW55379 V16111.MARC.IBOV.Bos
gb_est8:AA521753	+	85.00	172.93	1.63	590	AA521753 V151507.R1.Barstead.m
gb_est7:BB144606	+	79.00	165.41	4.28	327	BB144606 BB144606 RIKEN.full.1e
gb_est83:BF127246	+	76.50	151.11	26.81	947	BF127246 601650907R1.NIH.MGC.7
gb_est100:BG402468	+	70.50	137.34	156.72	1081	BG402468 60246617F1.NIH.MGC.7
gb_est57:BB133013	+	69.50	148.51	37.40	230	BB133013 BB133013 RIKEN.full.1e
gb_gss7:AA123031	-	69.00	141.27	94.75	474	AA123031 HS_3095.B1.B05.MF.CIT
gb_est72:BE232585	+	68.50	146.20	50.34	236	BE232585 137643.MARC.IPIG.Sus
gb_est67:BB516797	+	68.50	144.20	65.06	298	BB516797 BB516797 RIKEN.full.1e
gb_est21:AT500463	-	68.50	139.93	112.41	490	AT500463 tn97e01.x1.NCI.CGAP.Ut
gb_est86:BE307423	+	68.50	134.40	228.49	934	BE307423 601894071F1.NIH.MGC.17
gb_est90:BF588692	+	68.00	142.75	78.28	312	BF588692 7152e07.x1.NCI.CGAP.Bi
gb_est81:BE980038	+	68.00	135.78	191.53	704	BE980038 UI-M-CG0-bcr-e-10-0-UI
gb_est73:BE359717	+	67.50	140.34	106.72	366	BE359717 DGI_56.E11.92.A002.Dat
gb_est90:BF600042	+	66.50	135.92	188.15	480	BF600042 264492.MARC.3BOV.Bos
gb_est80:BE866656	+	66.50	131.72	322.25	783	BE866656 60179255F1.NIH.MGC.53
gb_est85:BF217720	+	66.50	126.92	596.61	1371	BF217720 601822545F1.NIH.MGC.5
gb_est82:BF029198	-	66.00	129.96	403.96	891	BF029198 601765414F1.NIH.MGC.53
gb_est80:BE899224	+	66.00	128.62	479.73	955	BE899224 601682045F1.NIH.MGC.9
gb_est38:AV434080	+	65.00	132.83	279.68	477	AV434080 AV434080 Porphyria.yezc
gb_est38:AV434616	+	65.00	132.42	294.54	500	AV434616 AV434616 Porphyria.yezc
gb_est38:AV432219	+	65.00	131.78	319.90	539	AV432219 AV432219 Porphyria.yezc
gb_est38:AV435766	+	65.00	131.62	326.43	549	AV435766 AV435766 Porphyria.yezc
gb_est38:AV436271	+	65.00	131.59	327.74	551	AV436271 AV436271 Porphyria.yezc
gb_est38:BG320355	+	65.00	128.02	518.35	836	BG320355 Zm03_05h01.R.Zm03.AAFC
gb_est85:BF242509	+	65.00	127.53	551.85	885	BF242509 601876084F1.NIH.MGC.55
gb_est82:BF042036	+	64.50	130.68	358.47	529	BF042036 BP250023A20F6.Soares
gb_est74:BE403953	+	64.50	130.89	368.17	542	BE403953 WHE0415.C10.F19ZS.Whea
gb_est39:AV624340	+	64.50	130.64	370.41	545	AV624340 AV624340 Chlamydomonas
gb_est99:BG309910	-	64.50	127.10	582.80	823	BG309910 HVSMC0014H07f.Hordeum
gb_est78:BE731965	-	64.50	125.64	702.98	976	BE731965 601568983F1.NIH.MGC.21

gb\_gss34:AZ922294 - 64.00 139.04 126.09 181 ! AZ922294 MRCot4H05 Sorghum b  
gb\_est96:BG063271 - 64.00 132.39 295.75 393 ! BG063271 H3005803-3 NIA Mous  
gb\_gss18:AQ883620 - 64.00 129.37 435.70 559 ! AQ883620 HS\_5478\_B2\_B02.T7A  
gb\_est99:BG310261 - 64.00 125.79 689.85 849 ! BG310261 HVSMC0016L07f.Hord  
seq\_name: gb\_est46:AW403229

seq\_documentation\_block:  
LOCUS AW403229 439 bp mRNA EST 16-FEB-2000  
DEFINITION UI-HF-BKO-abb-a-03-0-UI.r1 NIH\_MGC\_36 Homo sapiens cDNA clone  
IMAGE:3055397 5', mRNA sequence.

ACCESSION AW403229  
VERSION AW403229.1 GI:6922108  
FEATURES  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 439)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov

Eco RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: M.B. Soares Lab  
cDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Seq primer: M13 Forward.

FEATURES  
Location/Qualifiers  
1..439  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3055397"  
/clone\_lib="NIH\_MGC\_36"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MGC85"  
/lab\_host="DH10B (LT1)"  
/note="Vector: p7T3-Pac; Site\_1: NotI; Site\_2: Eco RI;  
Constructed from size fractionated cytoplasmic mRNA  
(0.5-1.5kb), directionally cloned. Cells provided by Louis  
M. Staudt, Ph.D. Library preparation by Maria de Fatima  
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 87 a 154 c 110 g 88 t  
ORIGIN

alignment\_scores:  
Quality: 178.00 Length: 32  
Ratio: 5.562 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-485-951-1 x AW403229 ..

Align seg 1/1 to: AW403229 from: 1 to: 439

1 AsnProArgThrValProValGlnProAlaPheSerThrValProPheSe 17  
|||||  
232 AACCCCGGACACCTCCCTGTTACAGCTGCTCTCCACGGTCCGTTCTC 281  
|||||  
17 rGlnProValCysPheProProArgProArgGlyArgGlnLys 32  
|||||  
282 CCAGCTGCTGTTTCCACCCAGGCCGCGGCGCAGACAAAA 327  
|||||

seq\_name: gb\_est53:AW86539  
seq\_documentation\_block:



REFERENCE 1 (bases 1 to 665)  
 AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.  
 TITLE The FAPESP/LICR Human Cancer Genome Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&t2=QV3-BT0537-221299-048-c06&t3=1999-12-22&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 23  
 High quality sequence stop: 664.

FEATURES  
 source  
 1..665  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BT0537"  
 /dev\_stage="Adult"  
 /note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 132 a 208 c 176 g 148 t 1 others  
 ORIGIN

alignment\_scores  
 Quality: 164.00 Length: 32  
 Ratio: 5.467 Gaps: 0  
 Percent Similarity: 93.750 Percent Identity: 93.750

alignment\_block:  
 US-09-485-951-1 x AW373817 ..

Align seg 1/1 to: AW373817 from: 1 to: 665

1 AsnProArgThrValProValGlnProAlaPheSerThrValProPheSe 17  
 |||||  
 485 AACCCCGGCACAGTCCCTGTCAGCCTGCTCTCCACGGTCCGTTCTC 534

17 rGlnProValCysPheProProArgProArgGlyArgArgGlnLys 32  
 |||||  
 535 CCAGCCTGTGTGTTCCACCCAGGCCAGGAGCGCAGACNAAAA 580

seq\_name: gb\_est29:AL559254

seq\_documentation\_block:  
 LOCUS AL559254 535 bp mRNA EST 16-FEB-2001  
 DEFINITION AL559254 LTI\_NFL008\_Tc2 Homo sapiens cDNA clone CS0DJ012YD19 5 prime, mRNA sequence.

ACCESSION AL559254  
 VERSION AL559254.1 GI:12904575  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 535)  
 AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).  
 FEATURES  
 source  
 1..535  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="CS0DJ012YD19"  
 /clone\_lib="LTI\_NFL008\_Tc2"  
 /sex="male"  
 /tissue\_type="T cells from T cell leukemia"  
 /note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL : <http://fulllength.invitrogen.com>"

BASE COUNT 97 a 183 c 126 g 89 t 40 others  
 ORIGIN

alignment\_scores  
 Quality: 152.00 Length: 32  
 Ratio: 5.429 Gaps: 0  
 Percent Similarity: 87.500 Percent Identity: 87.500

alignment\_block:  
 US-09-485-951-1 x AL559254 ..

Align seg 1/1 to: AL559254 from: 1 to: 535

1 AsnProArgThrValProValGlnProAlaPheSerThrValProPheSe 17  
 |||||  
 205 AACCCCGGCACAGTCCCTGTCAGCCTGCTCTCCACGGTCCGTTCTC 254

17 rGlnProValCysPheProProArgProArgGlyArgArgGlnLys 32  
 |||||  
 255 CCAGCCTGTGTGTTCCACCCAGGCCAGGAGCGCAGACNAAAA 300

seq\_name: gb\_est93:BF883419

seq\_documentation\_block:  
 LOCUS BF883419 404 bp mRNA EST 17-JAN-2001  
 DEFINITION QV3-ET0203-081200-525-a04 ET0203 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF883419  
 VERSION BF883419.1 GI:12273545  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 404)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001





Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH

BASE COUNT 219 a 251 c 237 g 201 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 106.00 Length: 27  
Ratio: 4.609 Gaps: 0  
Percent Similarity: 85.185 Percent Identity: 74.074

alignment\_block:

US-09-485-951-1 x BF181157 ..

Align seg 1/1 to: BF181157 from: 1 to: 909

6 ProValGlnProAlaPheSerThrValProPheSerGlnProValCysph 22  
|||||.....:|||||..... |||||..... ||  
488 CCTGTCCAGCCTGTCTCTCCAGTGCAGTTCCTCTCAGCCAGTCAGT 537

22 eProProArgProArgGlyArgGlnLys 32

|||||.....:|||||..... |||||..... ||  
538 CCCACGGACCCCTAAGGGCGCAACAGAAA 568

seq\_name: gb\_est89:BF533977

seq\_documentation\_block:

LOCUS BF533977 1165 bp mRNA EST 11-DEC-2000  
DEFINITION 602075139F1 NCI\_CGAP\_L19 Mus musculus cDNA clone IMAGE:4212273 5',  
mRNA sequence.  
ACCESSION BF533977  
VERSION BF533977.1 GI:11621340  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 1165)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM9782 row: m column: 10  
High quality sequence stop: 735.  
Location/Qualifiers  
1. 1165

FEATURES  
source  
1. 1165  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4212273"  
/clone\_lib="NCI\_CGAP\_L19"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: liver; Vector: pCMV-SPORT6; Site.1: NotI;  
Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.9 kb. Constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 331 a 320 c 286 g 228 t

ORIGIN

alignment\_scores:

Quality: 106.00 Length: 27  
Ratio: 4.609 Gaps: 0

Percent Similarity: 85.185 Percent Identity: 74.074

alignment\_block:

US-09-485-951-1 x BF533977 ..

Align seg 1/1 to: BF533977 from: 1 to: 1165

6 ProValGlnProAlaPheSerThrValProPheSerGlnProValCysph 22  
|||||.....:|||||..... |||||..... ||  
491 CCTGTCCAGCCTGTCTCTCCAGTGCAGTTCCTCTCAGCCAGTCAGT 540

22 eProProArgProArgGlyArgGlnLys 32

|||||.....:|||||..... |||||..... ||  
541 CCCACGGACCCCTAAGGGCGCAACAGAAA 571

seq\_name: gb\_est82:BF046316

seq\_documentation\_block:

LOCUS BF046316 323 bp mRNA EST 10-OCT-2000  
DEFINITION BP250002B10C9 Soares normalized bovine placenta cDNA  
clone BP250002B10C9 5', mRNA sequence.  
ACCESSION BF046316  
VERSION BF046316.1 GI:10763371  
KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 323)  
LEWIN,H.A., Soares,M.B., Pardinas,J., Liu,L. and Larson  
J.H.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

W. M. Keck Center for Comparative and Functional Genomics  
University of Illinois at Urbana-Champaign  
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL  
61801, USA  
Tel: 217 333 5998  
Fax: 217 244 5617  
Email: [h-lewin@uiuc.edu](mailto:h-lewin@uiuc.edu)  
Funding for cattle EST sequencing was provided by the USDA National  
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534  
to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED  
from Washington University Genome Center. Vector Trim: g;  
Cross\_match from Washington University Genome Center PHRAP suite.  
This sequence is vector free and at least 200 bp in length.

PCR Primers

FORWARD: TAATACGACTCCTACTATAGGG

BACKWARD: ATTAACCTCACTAAAG

Insert Length: 323 Std Error: 0.00

Plate: BP250002B10 row: C column: 9

Seq primer: ACGGATAACAATTTTCACACAGGA

High quality sequence stop: 323.

FEATURES

source

1. 323

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone="BP250002B10C9"

/clone\_lib="Soares normalized bovine placenta"

/sex="female"

/lab\_host="DH10B"

/note="Organ: placenta; Vector: pT7T3Pac; Site.1: EcoRI;

Site.2: NotI; The cDNA library was contributed by the

Soares laboratory and it was constructed and normalized

as described by Bonaldo, M.F., Lennon, G. and Soares,

M.B. (1996), Genome Research 6(9): 791-806."

BASE COUNT 59 a 132 c 66 g 63 t 3 others

ORIGIN

alignment\_scores:

Align seg 1/1 to: AM655379 from: 1 to: 539

```
1 AsnProArgThrValProValGlnProAlaPheSerThrValProPheSe 17
36 AACATCCGCGCATCCCAAGACCGCGCTGCCAAGGTGCAGTTCTC 85
17 rGlnProValCysPheProProArgProArgGlyArgArg 30
86 CCAGGTGTCTGTCTCCACCCAGGCCAGGGGGCGCAAA 125
```

seq_name: gb_est8:AA521753	seq_documentation_block:	590 bp	mRNA	EST
LOCUS	AA521753			
DEFINITION	vl15h07.r1 Barstead mouse proximal colon MPLRB6 Mu			
	clone IMAGE:903901 5' similar to SW:LEG4 RAT P3855			

mRNA sequence:  
 AA521753  
 ACCESSION

VERSION	AA521/33.1	GI:2202490
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE

Alloin M, Hillier I, Alloin M, Deleau M, Dietrich M, Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Murini; Mus (bases 1 to 590)

AUTHORS	Mallat,S., Hillier,D., Allen,M.M., Bowles,M., Dettliff,G., Giesel,S., Kucaba,T., Lacy,A.M., Le,M., Martin,J., Mochly,N., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Theising,B., Wylie,T., Lennon,G., Soares,B., Willson,Waterston,R.
TITLE	The WashU-HHMI Mouse EST Project

JOURNAL COMMENT

unpublished (1996)

Contact: Marra M/Mouse EST Project  
WASHU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63110  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL  
 IMAGE Consortium (info@image.llnl.gov) for further  
 MGI:524565  
 Seq primer: -28ml3 rev2 ET from Amersham  
 High quality sequence stop: 401.  
 Location/Qualifiers  
     1..590  
FEATURES  
     source

```

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:903901"
/clone_lib="Barstead mouse proximal colon"

```

```

/dev_stage="7 day juvenile"
/lab host="DH10B"

```

BASE COUNT  
ORIGIN

142 a 156 c 154 g 138 t

/note="Vector: pTT73D-Pac (Pharmacia) with  
polylinker; Site\_1: EcoRI; Site\_2: NotI;  
was primed with Not I - oligo(dT) primer:  
TGTACGAATCTGAAGTGGACGCGCCCTTTTTTTTTT  
3'); double-stranded cDNA was ligated to  
[AATTCGATCCTTG], digested with Not I and  
Not I and Eco RI sites of the modified p  
Library constructed by Bob Barstead. "

alignment_scores:			
Quality:	85.00	Length:	27
Ratio:	4.048	Gaps:	0
Percent Similarity:	77.778	Percent Identity:	62.963

alignment\_block:  
US-09-485-951-1 x AA521753

```
Align seg 1/1 to: AA521753 from: 1 to: 590

6 ProValGlnProAlaPheSerThrValProPheSerGlnProValCysPhe 22
||||| :|||:||||| 28-JUN-2000
508 CCTGTCCAGCGTGTCTCTCCACGAGTGTCTCTCAGCCAGTTTCAGTT 557

22 eProProArgProArgGlyArgArgGlnLys 32
||||| :|||:|||||
558 TCCCGCGGACCCCTAAGGCGCGCAACAGGAAA 588

seq_name: gb_est57:BB144606

seq_documentation_block: 327 bp mRNA EST
LOCUS BB144606 RIKEN full-length enriched, adult female vagina Mus
DEFINITION Musculus cDNA clone 9930026D03 3' similar to U81829 Mus musculus
calumenin mRNA, mRNA sequence.
BB144606
BB144606.1 GI:8799543
EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota: Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-resetc.riken.go.jp,
URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
Location/Qualifiers
1..327
/db_xref="taxon:10090"
/clone="9930026D03"
/clone_lib="RIKEN full-length enriched, adult female
vagina"
/sex="female"
/tissue_type="vagina"

/dev_stage="adult"
/lab_host="DH10B"
/note="Site:1: SalI; Site:2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 185.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGAGTTAATTAATTAATCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT 62 a 95 c 55 g 115 t
ORIGIN

alignment_scores:
Quality: 79.00 Length: 29
Ratio: 4.158 Gaps: 1
Percent Similarity: 65.517 Percent Identity: 55.172

alignment_block:
US-09-485-951-1 x BB144606 ..
Align seg 1/1 to: BB144606 from: 1 to: 327
2 ProArgThrValProVal.....GlnProAlaPheSe 12
||| :|||:|||||
1 CCCCCTCCACCCCGTGACTCACACTCACCACCACTGCTTCC 50

12 rThrValProPheSerGlnProValCysPhePro 24
|||||:|||||:|||||
51 TAGCATGTCATTTCTCGACCCGCGTCTCCACCC 87

seq_name: gb_est83:BF127246

seq_documentation_block: 941 bp mRNA EST
LOCUS BF127246 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4071744 3',
DEFINITION mRNA sequence.
ACCESSION BF127246
VERSION BF127246.1 GI:10966286
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM917 row: n column: 01
High quality sequence stop: 5.
FEATURES
source
Location/Qualifiers
1..941
/db_xref="taxon:9606"
/organism="Homo sapiens"
/db_xref="taxon:9606"/>
```

